

**ENSO ENSEMBLE PREDICTION AND PREDICTABILITY FOR
THE PAST 148 YEARS FROM 1856-2003**

by

Yanjie Cheng

M.Sc., Dalhousie University, 2006

M.Sc., Nanjing Institute of Meteorology (China), 1995

B.Sc., Nanjing Institute of Meteorology (China), 1992

Dissertation Submitted in Partial Fulfillment of
The Requirement for the Degree of
Doctor of Philosophy
In
Natural Resources and Environmental Studies

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

May 2010

© Yanjie Cheng, 2010



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-61141-8
Our file *Notre référence*
ISBN: 978-0-494-61141-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■
Canada

To my loved family
(献给我亲爱的家人)

Abstract

Several important issues of El Niño-Southern Oscillation (ENSO) predictability were studied using the latest version of the Zebiak-Cane model, singular vector (SV) analysis, ensemble hindcast, and information theory for the period of 148 years, e.g., the dominant factors controlling ENSO prediction skills, the useful precursors of forecast skill, ensemble construction and probabilistic verification.

More precisely, there are four main sections in this thesis. 1) A fully physically-based tangent linear model was constructed for the Zebiak-Cane model and a singular vector (SV) analysis for the 148 year (1856-2003) was performed. It was found that the leading SVs are less sensitive to initial conditions while singular values and final perturbation patterns exhibit a strong sensitivity to initial conditions. The dynamical diagnosis shows that the total linear and nonlinear heating terms play opposite roles in controlling the optimal perturbation growth. 2) Relationships between the singular values and actual prediction skill measures were investigated. At decadal/interdecadal time scales, an inverse relationship exists between the leading singular value (S1) and correlation-based skill measures whereas an in-phase relationship exists between the S1 and MSE-based skill measures. However, S1 is not a good predictor of prediction skill at shorter time scales and for individual predictions. An offsetting effect was found between linear and nonlinear perturbation growth rates, which have opposite contributions to the S1. 3) Ensemble and probabilistic ENSO predictions were performed for the 148 yrs. Four typical ensemble construction strategies were investigated. Results suggest that “reliability” is more sensitive to choice of ensemble construction strategy than “resolution”. The fourth strategy produces the most reliable and skillful ENSO

probabilistic prediction, benefiting from the contribution of the stochastic optimal winds and singular vector of SSTA. 4) Information and ensemble-based potential predictability measures are explored on multiple time scales. Relative entropy is better than predictive information (PI) and predictive power (PP) in quantifying the correlation-based prediction skill; whereas PI/PP is a better indicator in estimating mean square error (MSE)-based prediction skill.

Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor Dr. Youmin Tang and co-supervisor Dr. Peter Jackson for their guidance, mentoring, and constant encouragement during my study period in UNBC. Their keen and vigorous academic observation enlightens me not only in this thesis but also in my future study. Thanks Dr. Dake Chen, Dr. Xiaobing Zhou, and Dr. Ziwang Deng for their contributions to this work; they gave many helpful comments, suggestions, and discussions in my four manuscripts. Dr. Dake Chen provided the latest Zebiak-Cane model and the high quality initial dataset. Dr. Xiaobing Zhou helped me in learning the method and software to generate the tangent linear model. Dr. Ziwang helped me in learning the wavelet & cross-wavelet analysis methods. I would like to express my gratitude to the other committee members, Dr. Liang Chen, Dr. Jianbin Li, and Dr. Art Fredeen, for their suggestions in my thesis proposal and thesis research period. Thanks the discussions with Dr. Soon-IL An. Thanks Dr. B. P. Kirtman and anonymous reviewers for constructive comments on chapter 2 and 3.

This thesis greatly benefits from Youmin's courses, from which I learnt many helpful statistical methods and tools. Thanks Dr. Art Fredeen, Dr. Joselito M. Arocena, and Dr. Annie Booth for they opened my mind with interesting scientific topics of environmental sciences during my courses study.

I appreciate my two superiors, and Dr. Liang Chen, and UNBC graduate study program office Miss Bethany Haffner for their helps and encouragements in my application of the Natural Sciences and Engineering Research Council (NSERC) graduate scholarship.

I am grateful for the numerical computation supports of Dr. Jean Wang at the HPC Lab of UNBC. I owe my sincere gratitude to other group members Zhiyu Wang, Xiaoqin Yan, Jaison Ambadan, Dejian Yang, Dr. Yeqiang Shu, Waqar Younas, Siraj UL Islam, and Manoj Nambiar, for their friendships and discussions on research and careers. I also thank my friends Dr. Junhua Zhang, Dr. Bin Yu, Dr. Qiaobin Teng, Dr. Quanji Wu, Qiong Bai, Lin Bai, Jun Ma, Hongjun Mu, Li Huang, and all my friends in Canada.

My deepest thanks go to my beloved wife Kewei Xu, and my parents-in-law Jianmin Xu and Min He, and my parents Zhenxi Jing and Guiju Jia, for their loving considerations, encouragement, and especially their great works in taking care of my beloved daughter Lining Cheng during my 8yr's Master and PhD study period in Canada. Without their constant support and encouragement, I would not have finished my thesis.

This work is supported by Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) GR523, GR-7027, and the Graduate fellowship of the Natural Sciences and Engineering Research Council (NSERC) PGS D2-362539-2008.

CONTENTS

Abstract.....	i
Acknowledgement	iii
Contents	v
List of Tables	x
List of Figures.....	xi
Chapter 1: Introduction	1
1.1. ENSO predictability.....	1
1.2. Ensemble and probabilistic ENSO predictions	5
1.2.1. Ensemble Construction Strategies.....	6
1.2.2. Verification for Ensemble-Based Probabilistic Prediction.....	7
1.3. Information-based Measures of Potential Predictability.....	8
1.4. Objectives and Outline	12
Chapter 2: Further Analysis of Singular Vector and ENSO Predictability in the Lamont Model --- Part I: Singular Vector and the Control Factors.....	15
2.1.Introduction.....	16
2.2.Methods.....	19
2.2.1. ZC model LDEO5 version.....	19
2.2.2. Data and Model Initialization	20

2.2.3. Construction of the tangent linear model (TLM) of the ZC model.....	21
2.2.4. Data and Model Initialization.....	21
2.2.5. Theory of SV analysis.....	21
2.3.SV analysis over 148 years.....	25
2.3.1. Variations of the first SV and the final pattern.....	25
2.3.2. Variations of the singular value.....	29
2.4.Physical processes of perturbation growth in the ZC model.....	33
2.5.Conclusion.....	40
 Chapter 3: Further Analysis of Singular Vector and ENSO Predictability in the	
Lamont Model --- Part II: Singular value and predictability.....	58
3.1. Introduction.....	59
3.2. Methods.....	61
3.2.1 Metrics of Actual Prediction Skill.....	62
3.2.2 Cross-Wavelet Analyses.....	63
3.3. The Singular Value and ENSO Predictability.....	64
3.3.1 S1 – Predictability Relationship on the Decadal/Interdecadal Time Scale.....	64
3.3.2 S1 – Predictability Relationship on Interannual Time Scales.....	66
3.3.3 S1 - Predictability Relationship on All Time Scales.....	66

3.4. The Linear/Nonlinear Perturbation Growth Rates and the Actual Predictability.....	67
3.5. ENSO Signals, the Optimal Error Growth Rates, and Predictability.....	73
3.6. Conclusion.....	75
Chapter 4: Ensemble Construction and Verification of the Probabilistic ENSO Prediction in the LDEO5 Model.....	91
4.1. Introduction.....	92
4.2. Metrics for Ensemble Prediction Deterministic Skill.....	96
4.3. Strategies of Ensemble Construction.....	97
4.3.1. Perturbation of Initial Condition with Singular vector (SV) of SSTA.....	97
4.3.2. Realistic Stochastic Winds.....	98
4.3.3. Stochastic Optimal perturbation.....	98
4.3.4. Combination of Stochastic Optimal and Initial SSTA Perturbations.....	99
4.4. Verification Principles of Probabilistic Forecasts.....	100
4.4.1. Ensemble Spread and Error of Ensemble Mean.....	101
4.4.2. Reliability Diagram (RD).....	102
4.4.3. The Brier Score	102

4.4.4. The RPS Score.....	104
4.5. Results.....	106
4.5.1 Ensemble Spread.....	106
4.5.2 Reliability.....	109
4.5.3 Resolution.....	111
4.5.4. Overall Probabilistic Skill.....	112
4.6. Conclusion and discussion	113
Chapter 5: Information-Based and Ensemble-Based Potential Predictability Measures For ENSO Retrospective Forecast from 1856-2003.....	130
5.1.Introduction.....	131
5.2. The Strategy of Ensemble Construction.....	134
5.3. Prediction skill Metrics.....	136
5.3.1 Metrics of Actual Prediction Skill.....	136
5.3.2 Ensemble-based Measures of Potential Predictability.....	136
5.4. Characteristics of Information-based Measures of the LDEO5 Model.....	137
5.4.1 The general characteristics of RE, PI, and PP.....	137

5.4.2 Characteristic of Mutual Information.....	142
5.5. Relationship of Potential Predictability Measure and Actual Prediction Skill.....	145
5.5.1. The Relationship on the Decadal/Interdecadal Time Scale.....	145
5.5.2. The Relationship on Interannual Time Scales.....	147
5.5.3. The Relationship on All Time Scales.....	148
5.5.4. Cross-wavelet Analyses For Potential Measures and Actual Measures.....	149
5.6. The Control Factor of Potential Measures.....	151
5.7. Discussion and Summary.....	153
Chapter 6: Conclusion and Discussion.....	173
6.1. Thesis Summary.....	173
6.2. Discussion	177
References.....	179

List of Tables

Table 3.1 Correlation coefficients of potential predictability measures and actual predictability measures. The actual predictability measure is MSEIP, whereas the potential predictability measures include the leading singular value (S1), the linear perturbation growth rate (L1) and the nonlinear perturbation growth rate (N1). The LH and NH represent the linear heating and nonlinear heating items in SST governing equation, averaged over NINO3.4 region and over the optimal period of 12 months....	90
Table 5.1 Summarization of prediction skill measures used in this study. Prediction skill measures are as functions of either lead time (t) or initial time (i), or both.....	157
Table 5.2 Correlation coefficients between actual prediction skill and potential prediction skill over the 148 years (from January 1856 to December 2003).....	158

List of Figures

- Fig. 2.1 The first singular vector and the first final pattern of SSTA averaged in the 148 years,
a) the first singular vector of SSTA; b) the first final pattern of SSTA. (SV1 & FP
explain 32 % of the variance of R in the SVD analysis using Eq. 2.5) (Unit:
°C).....43
- Fig. 2.2 Same as Fig. 2.1 but perturbing both SSTA (°C) and wind field (m/s). a) The first
singular vector of SSTA; b) the first final pattern of SSTA; c) the first singular vector of
the wind field; d) the first final pattern of the wind field.....44
- Fig. 2.3 a) Spatial correlation coefficients between each first SV over 148-yr and the
averaged SV-1, the mean correlation coefficient is 0.85 (dash line); b) Accumulated
frequency against the spatial correlation coefficient. It indicates the fraction of SVs that
is smaller than the spatial correlation coefficient. For example, 20% of SV samples have
the spatial correlation less than the spatial correlation of 0.8..... 45
- Fig. 2.4 a) Composite SV1 of SSTA (°C) for (a) high spatial similarity cases (80% of total
SV1) and (b) low spatial similarity (20% of total SV1).46
- Fig. 2.5 the first SV of SSTA (°C) starting from the phase of a1) peak La Niña; b1) onset of
La Niña; c1) Neutral; d1) onset of El Niño; e1) Peak El Niño. The corresponding final
pattern after 6 months is shown in the right panel a2-e2. The averaged perturbation
growth rate S1 of each stage is marked in the title captions.....47
- Fig. 2.6 Same as Fig. 2.5 but for thermocline depth anomaly (H) (unit: m).....48

- Fig. 2.7 a) the SV and b) the final pattern (FP) averaged in the higher skill period (1876-1895 and 1976-1995) and c) SV and d) the final pattern in the lower skill period (1916-1955).....49
- Fig. 2.8 Seasonal variations of the first singular values S1 against the prediction target time at different lead time, 1, 3, 6 and 9 months (from bottom to top respectively) averaged over 148 years..... 50
- Fig. 2.9 Time series of the low-pass filtered (>24 months) a) the first singular value (S1) and b) the NINO3.4 index used for the wavelet analysis. Wavelet power spectrum of c) S1 and d) NINO3.4 using the Morlet wavelet. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The solid smooth curves in the left and right corners indicate the edge effects become important.....51
- Fig. 2.10 The cross-wavelet analysis for NINO3.4 SSTA index and the singular value S1. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left, and NINO3.4 SSTA index leading S1 by 90° pointing straight down.....52
- Fig. 2.11 Mean S1 (in solid star curve) as a function of the background ENSO phase. a) The bar curve is the NINO3.4 index of the composite background ENSO cycle (S1 is divided by 2); b) same as a) but using NINO3 index.....53
- Fig. 2.12 Final patterns for linear and nonlinear heating terms averaged in 148 years: a) the final pattern of horizontal linear heating (HL); b) the final pattern of horizontal nonlinear

heating; c) the final pattern of vertical linear heating (WL); d) the final pattern of vertical nonlinear heating(WN); e) the final pattern of total linear heating (HL+WL); f) the final pattern of total nonlinear heating(HN+WN). (Unit: °C).....54

Fig. 2.13 The seasonal variation of singular values for linear and nonlinear heating terms a) horizontal linear heating (HL, solid line), vertical linear heating (WL, dash line), horizontal nonlinear heating (HN, solid star line), and vertical nonlinear heating (WN, dash dot line). The original singular values (open circles); b) singular values for the linear heating (solid star line), nonlinear heating (dash dot line).....55

Fig. 2.14 EOF analyses for linear heating terms in Eq. (10). From top to bottom, Figures in the left panel are the EOF-1 spatial patterns of a) horizontal linear (HL), b) vertical linear (WL) and c) total linear heating (HL+WL). Their corresponding PCs are in the right panel. (Unit: °C/month).....56

Fig. 2.15 Same as Fig. 14 but EOF analyses results for nonlinear heating terms. Figures from top to bottom are a) horizontal nonlinear (HN), b) vertical nonlinear (WN) and c) total nonlinear heating (HN+WN).....57

Fig. 3.1 Interdecadal variations of a) anomaly correlation coefficient (R) and the singular value (S1); b) MSE and S1. A 25-yr running window was applied on all data at each month from Jan. 1856 to Dec. 2003. MSE measures are averaged over lead times of 1-12 months.....81

Fig. 3.2 a) Decadal/interdecadal variations of MSEIP and singular value (S1). A 10-yr low-pass FFT filter method was applied on these skill measures. b) Temporal variations of

the correlation coefficient between S1 and MSEIP over the 148 years, correlation coefficients were calculated in a 25-year running window.....	82
Fig. 3.3 The relationships between the singular value S1 and the actual predictability measures at interannual time scales using a 2-7-yr FFT filter. S1 against MSEIP.....	83
Fig. 3.4 Same as Fig. 3 but for all time scales without using an FFT filter.....	84
Fig. 3.5 The relationships between the linear (L1), nonlinear (N1), and the original (S1) singular values. a) L1 against S1. b) N1 against S1. c) S1-(L1+N1) against S1.....	85
Fig. 3.6 The relationships between the linear/nonlinear singular values and the prediction skill MSEIP at interannual time scales using a 2-7-yr FFT filter. a) L1 against MSEIP; b) N1 against MSEIP.....	86
Fig. 3.7 The cross-wavelet analysis for the singular values S1/L1/N1 and actual prediction skill MSEIP. a) S1 and MSEIP; b) L1 and MSEIP; c) N1 and MSEIP. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left, and singular values leading skills by 90° pointing straight down. (A 2-yr FFT filter was applied on all data before performing the cross-wavelet analyses).....	87
Fig. 3.8 The cross-wavelet analysis for ENSO signal (NINO3.4) and the singular values S1, L1, and N1.....	88

Fig. 3.9 Decadal/interdecadal variations of ENSO signal ($[NINO3.4]$; the solid line) and perturbation growth rates (dash lines): a) the linear perturbation grow rate (L1); b) the nonlinear perturbation growth rate (N1); c) the total perturbation growth rate S1. A 10-yr low-pass filter has been applied.....89

Fig. 4.1 Root mean square error (RMSE) of NINO3.4 SSTA for the control run (CTL, circle), ensemble mean ($RMSE_{EM}$, '+'), along with ensemble spread (SPRD, star), and climatological standard deviation of SSTA (dash-dot line) as a function of lead time (month), averaged over 1856-2003. a) RMSE & SPRD for SV1_sst method; b) UV_realstoc; c) SO1_wind method; d) SO1_wind+SV1_sst method.....117

Fig. 4.2 Same as Fig. 4.1 but for the first ensemble construction method SV1_sst, with a larger SSTA initial perturbation magnitude (1.5 times of that in the Fig. 4.1a).....118

Fig. 4.3 A sensitivity study for the SPRD by adjusting the strength of stochastic winds in the second ensemble construction method. The perturbation magnitude varies from 0.5 to 3.0 times of that of NCEP high-frequency winds.....119

Fig. 4.4 a) Anomaly correlation skill (R) from the control run (open line), ensemble mean forecast with different level of stochastic wind perturbation (0.5, 1, 2, 3 times) than the realistic NCEP winds. b) Same as a), but for RMSE.....120

Fig. 4.5 The 148-yr averaged leading mode of the stochastic optimal (SO) winds (m/s). This mode explains the 30-40% of the original variance of S in Eq.

(4.5).....121

Fig. 4.6 a1-a3) Reliability diagram (RD) for the first ensemble construction method: SV1_sst, at lead times of 6, 9, 12 months. In each plot, three reliability curves represent three ENSO categories: Warm (circle), neutral (star), and cold events (square). These are calculated based on 100-member ensemble hindcasts for all months over the 1856-2003. b1-b3) for the second method: UV_realstoc; c1-c3) SO1_wind; d1-d3) SO1_wind+SV1_sst.....122

Fig. 4.7 The reliability term B_{rel} of the Brier skill score as a function of lead time for the four ensemble prediction experiments at a) warm, b) cold, and c) neutral ENSO categories as functions of lead time (month). (Star: SV1_sst; dash-dot: UV_realstoc; square: SO1_wind; diamond: SO1_wind+SV1_sst). B_{rel} is negatively oriented.....123

Fig. 4.8 Analysis rank histograms for a small SPRD case (i.e. the second ensemble construction method UV_realstoc using the original high frequency winds) at different lead times 3, 6, 9, 12, 15, and 24 months. In a 100 ensemble members, the perfect percentage is 1.0% (dash line).....124

Fig. 4.9 Same as Fig. 4.8 but for a good/sufficient SPRD case, with a strong wind perturbation (3 times of the original high-frequency NCEP winds).....125

Fig. 4.10 a) The ranked probability score (RPS) as a function of lead time for three ensemble construction methods. b) Same as a) but for the ranked probability skill score (RPSS).

(Star: SV1_sst; dash-dot: UV_realstoc; square: SO1_wind; diamond: SO1_wind+SV1_sst).....	126
Fig. 4.11 Same as Fig. 4.7 but for the resolution term Bres of the Brier skill score. Bres is positively oriented.....	127
Fig.4.12 Same as Fig. 4.7 but for the Brier skill score (BSS). BSS is positively oriented...	128
Fig. 4.13 a) The ranked probability score (RPS) as a function of lead time for four ensemble construction methods. b) Same as a) but for the ranked probability skill score (RPSS).....	129
Fig. 5.1. The left panel: a1) Relative entropy (RE), b1) predictive information (PI), and c1) predictive power (PP) of the Niño3.4 SSTA index as a function of initial time of each prediction and lead time (months), from Jan. 1960 to Dec. 2003 for the LDEO5 model. The right panel: same as the left panel, but for lead time from 1-24-month.....	159
Fig. 5.2 a) Ensemble spread (ES) for time period from Jan. 1990 to Dec.1998 as a function of lead time (month) and initial time; b) same as a) but ensemble spread as a function of verification time.....	160
Fig. 5.3 a) Ensemble spread (°C) as a function of NINO3.4 SSTA index at lead times of 6, 9..., 24 month, respectively. ES is binned by SSTA at the initial time. b) Same as a) but at the target time.....	161
Fig. 5.4 a) Anomaly correlation of the LDEO5 ensemble mean forecasts of the monthly mean Nino3.4 SSTA over the 148-yr time period, as a function of target month	

(horizontal) and lead (vertical; in months); b) Same as a) but for RMSE; c) EM2; d) ES;
e) $\log(\text{RE})$; f) PP.....162

Fig. 5.5 Same as Fig. 5.4 but using a relative RMSE skill measure, defined by the RMSE over the RMSE of the climatological prediction. The climatological RMSE is as function of the calendar months.....163

Fig. 5.6 a) Variation of MI calculated by averaged RE and PI over all initial conditions as functions of lead times; b) MI from averaged RE and PI against the estimated MI from correlation skill using Eq. (5.4). c) Averaged RE and PI versus RMSE skill; d) Averaged RE and PP versus correlation skill.....164

Fig. 5.7 The estimated correlation skill calculated by Eq.(5.4) against the actual skill. in a) MI is obtained by averaged RE and averaged PI, respectively and in b) MI is obtained by the mean values of the averaged RE and averaged PI.....165

Fig. 5.8. Left panel: decadal/interdecadal variations of a1) RE and CIP; b1) PI and CIP; c1) PP and CIP. Right panel: decadal/interdecadal variations of a2) RE and MSEIP; b2) PI and MSEIP; c2) PP and MSEIP (A 25-yr running window has been applied for all data).....166

Fig. 5.9. Left panel: decadal/interdecadal variations of a1) RE and CIP; b1) PI and CIP; c1) PP and CIP. Right panel: decadal/interdecadal variations of a2) RE and MSEIP; b2) PI and MSEIP; c2) PP and MSEIP (A 10-yr low-pass FFT filter has been applied).....167

Fig. 5.10. Left panel: scatter plots of the individual correlation skill CIP against a1) RE; b1)

PI; c1) PP. Right panel: scatter plots of the individual MSE-based skill MSEIP against a2) RE; b2) PI; c2) PP (at interannual time scales using a 2-7-yr FFT filter).....168

Fig. 5.11 Left panel: the correlation skill CIP against a1) RE; b1) PI; c1) PP. Right panel: MSEIP against a2) RE; b2) PI; c2) PP.....169

Fig. 5.12. The cross-wavelet analysis for i potential measures and deterministic prediction skill measures. a) RE and CIP; b); RE and MSEIP; c) PI and CIP; d) PI and MSEIP; e) EM2 and CIP; f) EM2 and CIP; g) ER and CIP; h) ER and MSEIP. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left and information-based measures leading actual skill measures by 90° pointing straight down.170

Fig. 5.13. Temporal variations of relative entropy (RE; solid), the signal component (SC; solid thick line) and the dispersion component (DC; dash) of RE. (A 2-yr running mean method was applied on the data).....171

Fig. 5.14. Variations of a) RE, b) PI, and c) PP at the lead time of 6-month along with the time series of absolute value of the NINO3.4 index of SSTA. (A 2-yr running mean method was applied).....172

Chapter 1: Introduction

1.1. ENSO Predictability

El Niño/La Niña and the Southern Oscillation (ENSO), a coupled atmosphere-ocean interaction in the tropical Pacific Ocean, is the strongest interannual variability in the climate system. It happens in the tropical Pacific Ocean with a time period of 2-7 years and has global climatic, ecological, and social impacts. ENSO influences the mid-latitude regions through teleconnections and through ocean current anomalies. Significant impacts of ENSO on Canadian natural resources and the environment have been documented in a variety of areas including water resources, agriculture, forestry, fisheries, power utilities, coastal zones and other climate sensitive sectors of the Canadian economy (e.g., Hsieh et. al., 1999; 2003). For example, during El Nino, temperatures in the BC interior, especially in winter, are above normal and summer precipitation is typically below normal. Mountain pine beetle and forest fires are the two major natural disturbance agents in interior forests. Warm winter is favorable to mountain pine beetle survival and may lead to increases in lodgepole pine mortality. At the same time fire risk increases under warm and dry summer conditions.

ENSO Predictability is referred to as the extent to which it is possible to predict ENSO events. Generally, there are two sources that limit ENSO predictability: (i) uncertainty in initial conditions, and the chaotic behavior of the nonlinear dynamics of the coupled system (e.g., Jin et al. 1994; Chen et al. 2004); and (ii) atmospheric noise (i.e. weather events) and other high-frequency variations such as westerly wind bursts and the Madden-Julian oscillation (e.g., Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Vecchi and

Harrison 2003; Moore et al. 2006; Gebbie et al. 2007). In addition, model errors in the parameterizations of physical and dynamical processes also have impacts on the predictability. The first factor is referred to as the first kind of predictability, which is related to the nonlinear interactions and related instabilities within a chaotic system. And the second kind of predictability depends on the boundary conditions relevant for the system, such as the external atmospheric forcing. Some studies have suggested that the model-based prediction of ENSO depends more on the initial conditions than on unpredictable atmospheric noise (i.e., Tang and Hsieh 2003; Chen et al. 2004).

Significant progress has been made in understanding and predicting ENSO over the past decades. Many models with different levels of complexity such as simple models, intermediate coupled model, hybrid models and fully coupled general circulation models (GCM), have reached a correlation skill of 0.5 for predictions of 6-12 months or longer. However, some important issues still remain and are challenging to the ENSO community, including some important issues: i) Identifying the optimal growth of initial errors; ii) Estimating the prediction uncertainty; and iii) Seeking good measures of potential predictability that do not make use of observations, by which the degree of confidence that can be placed in an individual forecast can be assessed.

The first kind of predictability issue is inherent to the nature of ENSO prediction, the future evolution of the system depends critically on the initial state from which it started. A widely used strategy in studying the initial error growth is through the singular vector (SV) analysis, a mathematical method to measure the optimal error growth of nonlinear dynamic systems (an introduction to the SV analysis method will be given in section 2.3). The earliest

work using SV to explore the atmospheric prediction error growth due to uncertainties in initial conditions was documented in Lorenz (1965). In recent years, a number of models were used to explore the initial error growth of ENSO prediction using the SV analysis method. Chen et al. (1997) used the Battisti (1988) version of the Zebiak-Cane (ZC) model to calculate the SV and found that the optimal perturbation pattern consists of an east-west dipole in the entire tropical Pacific basin superimposed on a north-south dipole in the eastern tropical Pacific. Xue (1997a, b) constructed a forward tangent linear model for the ZC model using a Markov model and multi-variable EOF method performed on the reduced model physical space. Their SV spatial distribution was similar to that in Chen et al. (1997). Fan et al. (2000), using a different intermediate complexity coupled model, found that optimal error growth depends critically on the seasonal cycle and ENSO phase as well as the lead time of prediction. Tang (2006) studied the ENSO Predictability using a fully coupled GCM model and discussed some deficiencies in the GCM and their possible influences on SV growth.

A crucial component of any prediction system is the ability to estimate the predictive skill of a forecast so that the uncertainty of an individual forecast can be quantitatively estimated practically. This issue is often studied by using ensemble predictions, i.e., repeating a prediction many times, each time perturbing the initial conditions of a forecast model. A review of ensemble construction method is presented in section 1.2. Through ensemble prediction, the shape of the forecast probability density function (PDF) that describes the prediction uncertainty can be estimated. Under the assumption of a Gaussian process, a PDF can be characterized by its mean and variance, i.e. ensemble mean and ensemble spread.

Another important task in predictability study is to seek good measures of potential

prediction skill measures that do not make use of observations and actual prediction skills that are evaluated using observations, by which the degree of confidence that can be placed in an individual forecast can be assessed. Traditionally, the ensemble mean, ensemble spread and the ensemble ratio of signal-to-noise (ensemble mean over ensemble spread) are widely used as the measures of potential predictability to estimate the predictive skill a priori (e.g., Buizza and Palmer 1998; Moore and Kleeman 1998; Scherrer et al. 2004; Tang et al. 2008a). However, these ensemble measures have often met with challenges and limitations (Tang et al. 2005; Tang et al. 2007; Tang et al. 2008a, 2008b). In recent years, new ideas from information theory have made their appearance to examine ENSO and seasonal climate predictability, and many information-based measures have been used to quantify the predictability, such as information entropy, relative entropy, predictive information, mutual information (Schneider and Griffies 1999; Kleeman 2002, 2008; Tippett et al. 2004; Tang et al. 2005, 2008b; DelSole 2004, 2005, 2007, 2008).

While these issues have been addressed and studied for some years, all these aforementioned studies focused on a period of 20-30 years, so that the period available to test predictability covers rather few ENSO cycles (typically 10 or less), which precludes statistically robust conclusions. Chen et al. (2004) used KAPLAN SSTA reanalysis data and the LDEO5 version of the ZC model (LDEO5 hereafter) to perform a 148 years hindcast between 1856 and 2003. They successfully predicted almost all prominent El Niño events during this period at lead times of up to two years. Tang et al (2008b) further analyzed the interdecadal variation in ENSO prediction skill from 1881-2000 using multiple models. These retrospective ENSO predictions allow us to achieve a robust and stable study of

statistical predictability of ENSO. In this thesis, ENSO predictability will be explored in a period of over 100 years. Focus will be placed on exploring the optimal error growth of ENSO prediction and potential predictability, as well their relationship and decadal/interdecadal variations. With long-term SV analysis and corresponding retrospective ENSO ensemble prediction, it is expected that some new findings and understanding in ENSO predictability can be made.

1.2. Ensemble and probabilistic ENSO predictions

Ensemble forecasting has been widely used to explore the uncertainty of weather and climate predictions. Compared with a deterministic (single run) forecast, an ensemble forecast has many advantages. First, ensemble averaging acts as a nonlinear filter; it removes less predictable parts, and keeps more predictable features among the ensemble members (e.g., Leith 1974). A properly designed ensemble has higher skill than that of individual ensemble members in a statistical sense (Toth and Kalnay 1997). Second, ensemble prediction provides a practical tool for estimating the possible uncertainties in a forecast system. Ensemble forecasts can provide additional information, such as the probability density function (PDF) of a forecast, ensemble-based potential skill measures (i.e. ensemble mean, ensemble spread, and ensemble ratio), and probabilistic skill measures, which are useful in decision making. It is shown that probability forecasts have greater potential economic value than corresponding single deterministic forecasts with uncertain accuracy (e.g., Palmer 2000).

1.2.1. Ensemble Construction Strategies

Generally two kinds of strategies are used to produce optimal perturbations for ensemble-based ENSO predictability studies: i) perturbation of the initial conditions; ii) perturbations in the stochastic atmospheric noise through the whole forecast period. In addition, considering that model errors exist in physical/dynamical parameterizations, perturbation can be applied on model parameters, or using a multiple model ensembles approach (e.g., Kirtman and Min 2009). The first kind of strategy was often used by SV analysis (e.g., Lorenz 1965; Chen et al. 1997; Xue et al. 1997a, b; Battisti 1988; Fan et al. 2000; Cai et al. 2003, Tang et al. 2006) whereas the latter was performed in the framework of the stochastic optimal theory (e.g., Kleeman and Moore 1997, Moore and Kleeman 1998, 1999; Tang et al. 2005). Significant progress has been made in using these optimal perturbations to study ENSO predictability as cited above. However these previous studies mainly focused on the optimal error growth of ENSO deterministic predictions. The impact of perturbation construction on the ensemble probabilistic predictions has not been well addressed, especially using long-term retrospective ensemble predictions over periods as long as 100 years. In this study, we will explore this issue using SV-based perturbation methods. So far, the SV method itself has not been well examined in the framework of ENSO ensemble probabilistic prediction. One reason is that the SV analysis needs a tangent linear model (TLM), which is often technically difficult to produce. Another reason is the lack of long-term forcing data for initializing predictions, so that previous retrospective predictions were limited to a short period of 20-40 years, with a rather limited number of ENSO cycles. This may preclude statistically robust conclusions. In this thesis, a fully physically-based TLM

will be constructed for the LDEO5 model, and singular vector analyses will be performed for the 148 year period from 1856-2003. The SV analysis will make it possible to construct ensemble predictions with the LDEO5 model, so that the shape of the forecast probability density function (PDF) that describes the prediction uncertainty can be estimated, and the probabilistic nature of ENSO predictability can be explored.

Another issue is the role of stochastic atmospheric noise in ensemble ENSO predictions. It has been well recognized that stochastic atmospheric forcing associated with synoptic-to-intra-seasonal variability is critical in forming, developing and maintaining ENSO cycles (e.g., Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Eckert and Latif 1997; Blanke et al. 1997; Kirtman and Schopf 1998; Moore and Kleeman 1999; Thompson and Battisti 2000; Fluegel et al. 2004; Moore et al. 2006; Philip and van Oldenborgh 2009; Eisenman et al. 2005; Gebbie et al. 2007; Tziperman and Yu 2007; Zavala-Garay et al. 2003; Perez et al. 2005; Zhang et al. 2008). These studies consider that the high-frequency synoptic-scale atmospheric motion (i.e. weather events) and other high-frequency variations such as westerly wind bursts and the Madden–Julian oscillation (MJO) provide stochastic forcing to the ENSO modes and hence acting as a limit to the predictability. However, it is not very clear so far how the stochastic atmospheric noise impacts ENSO probabilistic predictions.

1.2.2. Verification for Ensemble-Based Probabilistic Prediction

An important task associated with ensemble construction is to evaluate an ensemble-based probabilistic prediction system by probabilistic verification methods, from which the

performance of the prediction system and the ensemble construction method can be quantitatively evaluated. Probabilistic verification is known as an important complement to deterministic verification, which provides a useful and quantitative way to measure uncertainty (Palmer 2000; Kirtman 2003). In contrast with the traditional prediction skill measures such as anomaly correlation (R) skill and root mean squared error (RMSE) skill, the verification of an ensemble-based probabilistic forecast system focuses on measuring two properties: reliability and resolution, which are the two most important characteristics of a probabilistic forecast system (Toth et al. 2003).

An introduction of these properties and probabilistic verification methods will be described in Section 4.4.

1.3. Information-based Measures of Potential Predictability

We now give a review of information-based measures of potential predictability; for further details, consult DelSole (2004) and Tang et al. (2008a). Information-based potential predictability measures the difference between two probability distributions: the forecast distribution $p(v | \Theta)$ and climatological distribution $q(v)$.

$$p(v | \Theta) = \int p(v | i) p(i | \Theta) di \quad (1.9)$$

Here, conditional probability $p(A|B)$ denotes the probability of A event when B event has happened. i , Θ , and v denotes the initial state, the corresponding observation (at initial time t), and forecast (at time $t+\Delta t$) respectively. Eq. (1.9) means that the forecast distribution $p(v | \Theta)$ can be theoretically obtained by the initial analysis probability $p(i | \Theta)$ and the

transition probability $p(v | i)$ of a perfect model system.

The climatological distribution $q(v)$ can be obtained by the long term forecast run (average) or observations. If the variable v is completely unpredictable, the forecast and climatological distributions will be identical, i.e. $p(v | \Theta) = q(v)$.

Entropy is a measure of dispersion level (e.g. uncertainty). The entropy of a continuous distribution $p(x)$ is defined as

$$H(x) = - \int p(x) \ln p(x) dx, \quad (1.10)$$

where the integral is understood to be a multiple integral over the domain of x . The larger entropy is associated with smaller probability and larger uncertainty.

A natural measure of predictability is the predictive information (PI), defined as the difference between the entropy of the climatological and forecast distributions:

$$PI = H(v) - H(v | \Theta) \quad (1.11)$$

Consider (1.10), then

$$PI = - \int q(v) \ln[q(v)] dv + \int p(v | \Theta) \ln[p(v | \Theta)] dv. \quad (1.12)$$

The first term on the rhs of Eq. (1.12) denotes the entropy of the prior distribution $q(v)$ (climatological distribution), measuring the uncertainty of a prior time when no extra information is provided from observation or model; whereas the second term represents the

entropy of the posterior distribution $p(v|\Theta)$ (forecast distribution), measuring the uncertainty after the observation and associated prediction becomes available. Thus a large PI indicates that the posterior uncertainty will decrease because of useful information being provided by a prediction (e.g., the larger $p(v|\Theta)$ the smaller uncertainty), that is, the prediction is likely to be reliable (Tang et. al. 2008c).

An alternative measure of the difference between two distributions is relative entropy (RE),

$$RE = \int p(v|\Theta) \ln \left(\frac{p(v|\Theta)}{q(v)} \right) dv \quad (1.13)$$

where q denotes the climatological distribution, and p is that for the prediction.

In the case where the PDFs are Gaussian distribution, which is a good approximation in many practical cases (including ENSO prediction), the relative entropy may be calculated exactly in terms of the predictive and climatological variances, and the difference between their means. The resulting analytical expression for the relative entropy R is given by (Kleeman 2002, Tang et al. 2008c):

$$RE = \frac{1}{2} \left\{ \ln \left[\frac{\det(\sigma_q^2)}{\det(\sigma_p^2)} \right] + \text{tr} [\sigma_p^2 (\sigma_q^2)^{-1}] + (\mu_p - \mu_q)^T (\sigma_q^2)^{-1} (\mu_p - \mu_q) - n \right\} \quad (1.14)$$

where, q and p are the climatological and predictive covariance matrices respectively; μ_q and μ_p are the climatological and predictive mean state vectors of the system, and n is the number of degree of freedom. R is composed of two components: (i) a reduction in

climatological uncertainty by the prediction [the first two terms plus the last term on the right-hand side of (1.14)] and (ii) a difference in the predictive and climatological means [the third term on the rhs of (1.14)]. These components can be interpreted respectively as the dispersion and signal components of the utility of a prediction (Kleeman, 2002). A large value of R indicates that more information that is different from the climatological distribution is being supplied by the prediction, which could be interpreted as making it more reliable (Tang et al. 2008a).

For a Gaussian distribution, a univariate state vector with a climatological mean of zero, the covariance matrices are scalar variances in (1.14). R , PI , and predictive power (PP) can be simplified as (DelSole 2004):

$$PI = \frac{1}{2} \ln \left(\frac{\sigma_q^2}{\sigma_p^2} \right) \quad (1.15)$$

$$RE = \frac{1}{2} \left[\underbrace{\ln \left(\frac{\sigma_q^2}{\sigma_p^2} \right)}_{\text{Dispersion}} + \underbrace{\frac{\sigma_p^2}{\sigma_q^2} - 1 + \frac{EM^2}{\sigma_q^2}}_{\text{Signal}} \right] = PI + \frac{1}{2} \left[\frac{\sigma_p^2}{\sigma_q^2} - 1 + \frac{EM^2}{\sigma_q^2} \right] \quad (1.16)$$

$$PP = 1 - \left(\frac{\sigma_p^2}{\sigma_q^2} \right)^{1/2} = 1 - \frac{\sigma_p}{\sigma_q} \quad (1.17)$$

A key difference between relative entropy (RE) and predictive information PI is that relative entropy RE vanishes if and only if the forecast and climatological distributions are identical (i.e., same mean and spread). Remarkably, predictive information and relative

entropy are invariant with respect to linear transformations of the state. The average of relative entropy and the average of predictive information have precisely the same value when averaged over all observations. This quantity is known as mutual information (MI) (DeSole 2004).

1.4. Objectives and Outline

The thesis study is being carried out through a seminal ENSO prediction model, i.e., Zebiak-Cane (ZC) model. For long retrospective predictions, a historic sea surface temperature of the past 148 years from 1856-2003 has been assimilated into the coupled model. An ensemble strategy that has been widely used to explore the uncertainty of weather and climate prediction will be used for quantitatively measure the information provided by predictions. To construct ensemble predictions, SV-based optimal perturbation methods will be used based on the singular vector (SV) analysis of the 148 years. A newly developed set of theoretical tools will be used to explore some essential issues related to ENSO prediction and predictability including the dominant precursors of forecast skill and the degree of confidence that can be placed in an individual forecast. Emphasis will be placed on using long-term retrospective ENSO prediction to derive stable and robust conclusion and findings. Since current studies of ENSO predictability usually use hindcasts of 20-30 years, the period available to test predictability covers rather few ENSO cycles (typically 10 or less) precluding statistically robust conclusions.

The long-term objective of my research is to significantly improve our capability in predicting ENSO/climate variability and in using ENSO/climate prediction. The short-term

specific objectives of this research are:

- i) Construct a tangent linear model TLM of the original ZC model for SV analysis.
- ii) Perform SV analysis over the 148 years. The main characteristics of the leading SV, final pattern, and perturbation growth rate will be investigated for the 148 years from 1856 to 2003. The controlling factors and mechanisms of perturbation growth rates will be discussed.
- iii) Explore long-term variability of singular values and its relationship with actual predictability measures. The relationship has very practical significance and offers a practical means of estimating the potential predictability and the confidence level of an individual prediction. The relationship between singular value and real predictability has not been addressed in previous studies of ENSO predictability due to the lack of sufficiently long retrospective prediction and corresponding SV analysis.
- iv) Carry out ensemble retrospective forecast of ENSO for 148 years from 1856-2003. A reliable and high resolution prediction system is fundamental in making ensemble and probabilistic ENSO predictions, and also important in investigating ENSO predictability with potential predictability measures. A perfect model will be convenient in applying the information theory to obtain information-based potential predictability measures. Thus, toward this objective, several ensemble construction strategies will be discussed and their predictions will be verified by probabilistic verification methods.
- v) Use information theory to derive robust measures of the uncertainty of ENSO predictions, and to identify the mechanisms responsible for the uncertainties of ENSO predictions, in order to find some good information-based and ensemble-based potential predictability

measures of ENSO prediction and exploring their relationships with actual prediction skill measures.

This thesis comprises four main chapters 2-5 to achieve the aforementioned goals. In Chapter 2, a TLM model is constructed for the ZC model, and SV analyses are performed for the 148 years and the controlling factors of perturbation growth are discussed. The third chapter explores the relationships of singular value and actual prediction skill. Emphasis is on discussing the scale-dependent features of the S1-skill relationship and explaining the good relationships between linear/nonlinear growth rate and actual prediction skills. In Chapter four, several typical ensemble construction methods are applied in ensemble predictions. Using probabilistic verification methods, the best method from the SV1_sst+SO1_winds was identified, indicating the important role of stochastic optimal perturbation at long lead times. Chapter 5 explores the relationships of potential predictability measures and actual prediction skill on multiple time scales. The controlling factors leading to a good potential measure are given in this chapter. Chapter 6 presents a summary of the principal results obtained in the thesis and some suggestions for future work.

**Chapter 2: Further Analysis of Singular Vector and ENSO Predictability
in the Lamont Model --- Part I: Singular Vector and the Control Factors**

Cheng Y, Tang Y, Zhou X, Jackson P, Chen D (2009) Further analysis of singular vector and ENSO predictability in the Lamont model—Part I: singular vector and the control factors. *Climate Dynamics*. DOI 10.1007/s00382-009-0595-7.

Published version is available at:

<http://www.springerlink.com/content/p328671887136l08/>

2.1. Introduction

ENSO is the strongest interannual variability in the global climate system. It happens in the tropical Pacific Ocean with a period of 2-7 years and has world-wide climatic, ecological, and social impacts. Significant progress has been made in understanding and predicting ENSO over the past few decades. At present, there are many ENSO prediction models with differing levels of complexity, including intermediate coupled models, hybrid coupled models and fully coupled general circulation models (GCM). When measured by the anomaly correlation between the predicted and observed sea surface temperature anomalies (SSTA) in the eastern Pacific, these models generally have prediction skills as measured by the correlation over 0.5 for lead times of 6-12 months (Latif 1998; Kirtman et al. 2002; Chen and Cane 2008). However, some important issues still remain unsolved such as the relationship between potential predictability and the actual prediction skill and the control factors of predictability.

A widely used strategy in studying initial perturbation growth is through singular vector (SV) analysis, a method to describe optimal perturbation growth. The earliest work using SV analysis to explore the growth of initial errors was documented in Lorenz (1965). In recent years, a number of models have been used to explore optimal perturbation growth of ENSO predictions using SV analysis. Chen et al. (1997) used the Battisti (1988) version of Zebiak-Cane (ZC) model to calculate the SV and found that the optimal perturbation pattern consists of an east-west dipole in the entire tropical Pacific basin superimposed on a north-south dipole in the eastern tropical Pacific. Xue et al. (1997a, b) constructed a tangent linear model in an EOF-reduced space for the ZC model via the Markov method. Their SV spatial distribution was similar to that of Chen et al. (1997). Fan et al. (2000), using a different intermediate complexity

coupled model, found that the optimal perturbation growth depends critically on the seasonal cycle and ENSO phase as well as the prediction lead time. Tang et al. (2006) studied ENSO potential predictability using a fully coupled GCM and discussed some deficiencies in the GCM and their possible influences on SV growth. Zhou et al. (2008) explored the impact of atmospheric nonlinearity on the optimal perturbation growth by comparing SVs of two ENSO models that have the same oceanic model coupled, respectively, to a linear and a nonlinear statistical atmospheric model.

However, there are still challenging issues concerning optimal perturbation growth that warrant further investigation. First, all of the above studies focused on a period of only 20-40 years, with a rather limited number of ENSO cycles, basically precluding statistically robust conclusions. A longer-term SV analysis would result in more robust ensemble feature of SV. Second, it has been well recognized that the actual predictability of ENSO has striking decadal/interdecadal variations (e.g., Chen et al. 2004; Tang et al. 2008a). One might be able to shed light on the mechanism of decadal/interdecadal variation in ENSO predictability by exploring decadal/interdecadal variation of the optimal perturbation growth by SV analysis. Obviously, the SV analysis for only a 20-40 year period, as performed previously, is unable to achieve this goal. Third, it has been of great interest to identify the sources and processes that limit the predictability of ENSO. Nonlinearity and stochastic noise are generally thought to be two most important factors limiting ENSO predictability. One effective method to explore the importance of nonlinearity in ENSO predictability might be to examine the relative roles that linear and nonlinear processes play in optimal perturbation growth, which has not been well addressed in previous studies. Finally, the relationship between optimal perturbation growth and the actual model prediction skill, i.e., between the potential predictability and actual

predictability, should be examined under a framework of statistically robust analysis.

Thus, further SV analysis is required to more fully understand optimal perturbation growth and ENSO predictability. In this first part of a two paper study of ENSO predictability, the first three challenges discussed above are addressed. In part two of the study, we will focus on actual model prediction skills and their relationship to optimal perturbation growth over a long-term period, which will provide insights on mechanisms of ENSO predictability. Recently, Chen et al. (2004) used KAPLAN sea surface temperature anomaly (SSTA) reanalysis data and the ZC model (LDEO5 version) to perform a 148 year hindcast experiment for the period of 1856-2003. They successfully predicted all of the prominent El Niño events during this period at lead times of up to two years, with the SST being the only data used for model initialization. Tang et al (2008a) further analyzed the interdecadal variation in ENSO prediction skill from 1881-2000 using multiple models. These retrospective ENSO predictions allow us to achieve a robust and stable study of statistical predictability of ENSO.

In the present paper, we perform SV analysis for the ZC model version LDEO5, from 1856-2003 using a newly constructed tangent linear model (TLM), then explore ENSO predictability using SV analysis. To our knowledge, this study is the first attempt to explore optimal perturbation growth of ENSO predictions by SV analysis for a period over 100 years. Emphasis will be placed on the first three aforementioned issues, in particular, investigating possible control factors and mechanisms responsible for variations in the SV. Section 2.2 briefly introduces the LDEO5 model, the construction of the tangent linear model, and the SV method. Section 2.3 presents the optimal perturbation growth pattern and perturbation growth rate by SV analysis. In section 2.4, the variability of SSTA is dynamically diagnosed and the dominant

factors controlling the perturbation growth, i.e. nonlinear heating (NH) and linear heating (LH), are discussed, followed by a conclusion and discussion in section 2.5.

2.2 Methods

2.2.1. Zebiak-Cane Model LDEO5 Version

The model used in this study is the Zebiak and Cane model (Zebiak and Cane 1987; hereafter ZC), which has been widely applied for ENSO simulation and prediction. LDEO5 is the latest version of the ZC model (Chen et al. 2004). The atmosphere dynamics follows Gill (1980) using steady-state, linear shallow-water equations. The circulation is forced by a heating anomaly which depends on the SST anomaly and moisture convergence. The ocean dynamics uses the reduced-gravity model, and ocean currents were generated by spinning up the model with monthly wind. The thermodynamics describe the SST anomaly and heat flux change. The model time-step is 10 days. The spatial region is focused on the tropical Pacific Ocean (124 °E-80 °W; 28.75 °S - 28.75 °N). The grid for ocean dynamics is 2° longitude 0.5° latitude, and the grid for SST physics and the atmospheric model is 5.625° longitude 2° latitude.

The SSTA dataset used in this study is a reconstructed analysis data by Kaplan (1998) with the period from January 1856 to December 2003. It is only an oceanic dataset available for initializing long-term retrospective ENSO prediction over 100 years. With the initialization of the SSTA dataset, the LDEO5 model successfully predicted all of the prominent El Niño events during at lead times of up to two years, and achieved a good hindcast skill (e.g., Chen et al. 2004; Tang et al. 2008). Note that in the coupled initialization procedure of the LDEO forecast system, assimilated SST data are not simply putting a constraint on the ocean model with SST

observations; they translate to surface wind field and subsurface ocean memory.

There are two model output statistics (MOS) schemes to correct model bias in the LDEO5. One scheme is for SST, and the other is applied to thermocline depth and winds. Bias correction terms are given at each time step (Chen et al. 2000). With the two statistical bias correction schemes, the imbalance among those model variables (e.g., SST, thermocline depth, and winds) due to SST assimilation or perturbation of initial SST in the framework of ensemble can be expected to quickly adjust during the prediction period.

2.2.2. Data and Model Initialization

The SSTA dataset used in this study is the reconstructed analysis of Kaplan (1998) for the extended period of 1856–2003. It is the only the initial data in the current retrospective study, identical to that in Chen et al. (2004). Initialized with this monthly analysis, a forecast with lead times up to 24 months was made from each month of the 148-yr period. The same data set was also used to verify the model predictions. The skillful retrospective predictions initialized by this historic SST data from Jan. 1856-Dec. 2003 was shown in Chen et al. (2004) and Tang et al. (2008a). The reason is like that given in Chen et al. (2004) as follows. “The LDEO5 model has a higher predictive skill when multiple data sets—sea level, winds, SST—are used for initialization, and its skill decreases only slightly when assimilating only SST data. We have to rely on SST data here because tropical Pacific sea level observations are virtually non-existent before 1970, and historic wind information is sparse and poorly calibrated. Note that in the coupled initialization procedure of the LDEO forecast system, assimilated SST data are not simply putting a constraint on the ocean model with SST observations; they translate to surface wind field and subsurface ocean memory.”

There are two bias correction (MOS) schemes included in the LDEO5 model. One MOS scheme is for SSTA, the bias correction term is given at each time step based on regression multiple variables EOF as discussed in Chen et al. (2000). Another one is for thermocline depth, currents and winds. With these online statistical bias correction schemes, the balance of those variables can be achieved during the prediction period of the LDEO5 model.

2.2.3. Construction of the tangent linear model TLM for the ZC model

To study the evolution of initial error, the linearized operator L of original nonlinear model, i.e., the tangent linear model TLM should be required. In this study, the TAPENADE, an Automatic Differentiation Engine^{*}, was used to construct the TLM from the original ZC model.

To ensure the constructed TLM is correct, a test procedure was performed as below:

i) A small perturbation was added to the initial condition of TLM and the original model respectively. The final patterns from two models were compared after 6-month model integration. The results show that there is little difference between the two final patterns.

ii) The singular vectors derived from the TLM were compared with the SVs in Chen et al. (1997) and Xue et al. (1997a). Their similarity verifies the correctness of the TLM

2.2.4. Theory of Singular Vector Analysis

The evolution of a small perturbation X of the initial state vectors of a nonlinear dynamical model can be represented as:

^{*} www-sop.inria.fr/tropics/tapenade.html

$$\frac{\partial X}{\partial t} = LX \quad (2.1)$$

where L is the linearized operator of the nonlinear model. At time $t + \Delta t$, the solution to Eq. (2.1) is given by,

$$X(t + \Delta t) = R(t, \Delta t)X(t) \quad (2.2)$$

R , a function of time and the lead time, is often called the propagator and represents the perturbation growth matrices. From (1.1) and (1.2),

$$R(t, \Delta t) = \exp\left(\int_t^{t+\Delta t} L dt\right) \quad (2.3)$$

For the whole model domain, the amplitude of perturbation growth is defined as below,

$$\begin{aligned} A &= \frac{\|X(t + \Delta t)\|}{\|X(t)\|} = \frac{\langle X(t + \Delta t), X(t + \Delta t) \rangle^{1/2}}{\langle X(t), X(t) \rangle^{1/2}} \\ &= \frac{\langle RX(t), RX(t) \rangle^{1/2}}{\langle X(t), X(t) \rangle^{1/2}} = \frac{\langle X(t), R^* RX(t) \rangle^{1/2}}{\langle X(t), X(t) \rangle^{1/2}} \end{aligned} \quad (2.4)$$

Where $\langle \rangle$ denotes the inner product, R^* is the transpose of R . An L-square norm is used in Eq. (1.4). The eigenvector (E) of R^*R is the SV of R , representing the perturbation growth patterns. Thus the SV can be obtained by two methods: the empirical orthogonal function (EOF) analysis for R^*R matrix or singular value decomposition (SVD) analysis. In this study, we use the second method,

$$R = F\Lambda E^* \quad (2.5)$$

Where Λ is a real, positive, diagonal matrix; E and F are orthonormal (unitary) matrices. The columns of E and F are SVs and final patterns (FP). From (1.5), we can see the relationship between the first SV mode (E_1) and the first final pattern FP mode (F_1):

$$R(t, \Delta t)E_1 = \lambda_1 F_1 \quad (2.6)$$

λ_1 is the largest singular value in the Λ matrix, representing the amplitude (rate) of the optimal perturbation growth (E_1).

Generally, there are two approaches for SV analysis: a direct method and an indirect method. The direct method derives the linearized operator L in (2.1) and its adjoint operator from the original nonlinear model, i.e., the tangent linear model (TLM) and the adjoint model (AM), both being used for calculating the derivative and gradient of model state variables. The procedure of the direct method is to run the original model, TLM, and AM simultaneously together with an SVD (Singular value decomposition) algorithm. The application of the direct method can be found in the literature (e.g., Moore and Kleeman 1996; Li et al. 2005). The indirect method uses two steps to get the propagator (R) in (2.2). The first step is to integrate the original model from initial time to several months later (i.e. the optimal period) and to record the final state X_f' . In the second step, small perturbations, denoted by X_i , are added in the initial field of the original model and the original model runs grid by grid. The final state, denoted by X_f'' , is recorded. The perturbation growth during the optimal period, denoted by X_f is the difference between X_f'' and X_f' and the propagator R is thus $(X_f'' - X_f')X_i^{-1}$. The maximum possible perturbation growth is the first (largest) singular value of the propagator R . The initial and final patterns that accomplish this perturbation growth are the right and left singular vectors of R .

In this study, we propose a mixed algorithm for SV analysis, in which the TLM model was directly constructed from the original ZC model but only used for producing R . The advantage of this mixed algorithm is that it maintains the computational accuracy by using TLM and avoids the technical difficulty inherent in producing the AM model. In implementation, given a perturbation onto a model grid, the TLM model integrates forward once; so that the TLM model runs as many times as the number of model grids. The initial SSTA perturbation is 0.05°C , about 1% of the original SSTA amplitude. It should be noted that the SVs are not very sensitive to the amplitude of initial perturbations when the initial perturbation varies between $\pm 0.25^\circ\text{C}$ for SSTA, ± 2 m/s for zonal and meridional wind anomalies, and ± 2 m for thermocline depth anomaly (H). The total perturbation growth during the optimal period (X_f) is actually a final pattern responding to the initial perturbation (X_i). The relationship between the X_f and the X_i can be described by (2.7), i.e.

$$R(t, t + \Delta t)X_i = X_f \quad (2.7)$$

To avoid calculating the inverse matrix X_i^* , the initial perturbation is fixed at 0.05°C , thus X_i is a diagonal matrix with all diagonal elements equal to 0.05°C .

$$R(t, t + \Delta t) = \frac{X_f}{0.05} \quad (2.8)$$

Finally, from Eq. (2.5), we can find the SVs, final patterns, and singular values.

2.3. SV analysis over 148 years

2.3.1 Variations of the first SV and the final pattern

First, we only consider initial uncertainties in SST. The SV analysis is performed every month at the optimal period of 6 months (i.e., 6-month lead) for 1856-2003, using the TLM and SVD method, as discussed in section 2.2. In each SV analysis, the optimal perturbation growth pattern (the first singular vector, SV1), final pattern, and perturbation growth rates (singular values) are obtained. Fig. 2.1a and Fig. 2.1b show the averaged SV1 and the corresponding final pattern at 6-month lead time over 148 years. As can be seen, the SV1 is dominated by a west-east dipole in the tropical Pacific Ocean: one center located south of the equator in the eastern tropical Pacific Ocean and the other located in the central Pacific Ocean near 150°W (Fig. 2.1a). Such a dipole structure favorable for the perturbation growth is probably inherent in ENSO dynamics. For example, the zonal SSTA gradient at the equatorial eastern Pacific weakens local upwelling and intensifies the warm Kelvin waves propagating eastward according to the delayed oscillator theory (Suarez and Schopf 1988). The warm eastward propagating Kelvin waves bring warm waters to the eastern Pacific Ocean and further intensify the anomalies, finally leading to an El Niño-like pattern as shown in Fig. 2.1b. Fig. 2.1a and Fig. 2.1b are similar to that in the SV1 and FP of the Battisti coupled atmosphere-ocean model (Chen et al. 1997) and an older version of the ZC model (Xue et al. 1997a).

In a coupled ocean-atmosphere model, initial uncertainties may come from the atmosphere as well. To examine the sensitivity of the SV1 and the final pattern to uncertainties in the atmosphere, we repeated the above SV analysis but included perturbations on the initial conditions of both the SSTA and anomalous wind (zonal wind U and meridional wind V). The initial perturbation of winds is 0.05 m/s in Fig. 2.2c. The results show that the spatial

distributions of the new SV1 (Fig. 2.2a) and the final pattern of SSTA (Fig. 2.2b) are similar to those shown in Fig. 2.1a and Fig. 2.1b, indicating the SV1 and the final pattern of SSTA are mainly determined by the uncertainty in SST itself. This similarity is because the stochastic atmospheric noise is not included in the ZC model and uncertainties in winds are highly related to those in SST, thereby, they can be well represented by SST uncertainties. The adjustment of the atmosphere to ocean variables such as SST and upper ocean heat content is fast, making the atmosphere a “slave” to the ocean at monthly or longer time scales. Warm SST causes atmospheric convection, resulting in a convergence of mass in the atmosphere on both sides of the equator as shown in Fig. 2.2c, the SV1 of winds. Correspondingly, the final pattern of winds shows a strong association with El Niño. For example, large westerly wind anomalies prevail over the central equatorial Pacific. The close relationship between SST and the surface wind stress over the tropical Pacific has been documented in many studies. The tropical atmosphere responds to large-scale SST anomalies in a coherent and reproducible way; the tropical flow patterns, especially over the open ocean, are so strongly determined by the underlying SST that they show little sensitivity to changes in the initial conditions of the atmosphere (e.g., Stern and Miyakoda, 1995; Shukla 1998). Vialard et al. (2005) performed a series of ensemble forecasts by the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system using wind, SST perturbation and random perturbation to the atmosphere during the forecast, individually and collectively. Their results suggested that the uncertainties in SST determine the spread of ensemble forecasts during the first two months of the forecast, while perturbations of the wind stress or atmospheric internal variability alone underestimate the perturbation growth during the early months of the forecast. Therefore these results suggest that ENSO predictability depends more on initial conditions in SST than in atmospheric winds.

However, the air-sea coupled components are much deterministic than the uncoupled atmospheric noise, thus, to a certain extent, uncertainties are supposed to be originated from such noise. Because the atmospheric noise component is not fully considered in the ZC model, it leaves room to improve the ENSO predictability by including stochastic atmospheric noise. A more useful forecast strategy might be to perform ensemble predictions and evaluate the uncertainties of the forecast system and ENSO predictability using probabilistic methods (Chen and Cane, 2008). And the SV method is one of the widely used ensemble construction methods to generate the probabilistic weather forecasts.

It has been found in previous work that the SV1 is not sensitive to initial conditions in many models (i.e., Chen et al., 1997; Xue et al., 1997a). It is of interest to further explore the sensitivity of SV1 to initial conditions using a long-term analysis. To do this, we calculated spatial correlations between the 148-yr averaged SV1 and each individual SV1, which measures the similarity among individual SV1s. The result is shown in Fig. 2.3. For most cases (over 80%), the spatial correlation coefficients are over 0.80, with an overall average of 0.85 for all initial conditions (148×12 months). Even though the majorities (80%) of SV1s are similar, it is interesting to know the differences of initial patterns for those (20%) SV outliers from majority SVs. Composite maps of SV1 are made for those 80% and 20% cases, as shown in Fig. 2.4a and Fig. 2.4b, respectively. As can be seen in Fig. 2.4, the difference between the two SV1 patterns is small, only manifested in the equatorial western Pacific. The strong spatial similarity in Fig. 2.3 and small difference in Fig. 2.4 indicate that SV1 is indeed insensitive to initial conditions in the ZC model. A stable SV1 pattern will be useful in ensemble construction to improve the resolution of ensemble-based probabilistic forecasts.

SV1 is also insensitive to the background SST of the ENSO phase. Based on a threshold of $\pm 0.5^{\circ}\text{C}$ of NINO3.4 SSTA (SSTA over the region 5°S - 5°N , 120°W - 170°W), El Niño and La Niña events are defined when the threshold is met for a minimum of 5 consecutive months. The peak phase and the onset phase of La Niña are further defined by $\text{NINO3.4 SSTA} < -1.2^{\circ}\text{C}$ and $-0.5^{\circ}\text{C} < \text{NINO3.4 SSTA} < -1.0^{\circ}\text{C}$, respectively. The neutral ENSO state, onset of El Niño, and the peak El Niño phase are defined by $|\text{NINO3.4 SSTA}| < 0.5^{\circ}\text{C}$, $0.5^{\circ}\text{C} < \text{NINO3.4 SSTA} < 1.0^{\circ}\text{C}$, and $\text{NINO3.4 SSTA} > 1.5^{\circ}\text{C}$, respectively. For each stage, a composite SV1 and a corresponding final pattern over 148 years are presented in Fig. 2.5. All SV1s in different ENSO stages have a similar west-east dipole pattern in the equatorial Pacific and with very similar amplitude. The spatial coverage of final patterns, however, slightly varies with ENSO phases. As seen in Figs. 2.5b2-d2, at the onset and neutral ENSO stages, final patterns span over almost all the equatorial Pacific; whereas at peak ENSO stages final patterns shrink and are confined to the east side of the dateline. Fig. 2.6 shows the SV1 of thermocline depth anomaly (H) and their final patterns for different ENSO phases. Similar to SSTA, the leading SV mode of thermocline is not sensitive to ENSO background, as expected. After six months, the final patterns show some differences among ENSO phases, although the major features remain consistence, i.e., thermocline deepening in the east and shoaling in the west.

It is of interest to explore the variability of SV1 and final pattern at interdecadal time scales. Based on the prediction skill presented in Chen et al. (2004) and Tang et al. (2008), we selected two 40-yr SV1s and final patterns from the 148-yr SV1 results. The model forecast correlation skill in the 40-yr period of 1876-1895 and 1976-1995 was high; and another 40-yr period is 1916-1955, with a low correlation skill. It was found that the composite SV1 and final pattern in two high prediction skill periods are very similar to each other with the averages shown in

Fig. 2.7a and Fig. 2.7c. As expected, the SV1 of 1916-1955 shown in Fig. 2.7b is also very similar to Fig. 2.7a due to the fact that SV1 is not sensitive to initial conditions. In contrast to this time invariant feature of SV1, final pattern changes significantly between high and low prediction skill periods. As seen in Fig. 2.7, the final pattern has a weaker perturbation growth amplitude and a smaller spatial coverage in the high correlation skill period; final pattern in the low correlation skill period is more than twice as large as the final pattern in the high skill period. Therefore, there is an inverse relationship between the prediction skill of the model and the amplitude of final pattern on the interdecadal time scale.

2.3.2 Variations of the singular value

The first singular value (S1) represents the fastest perturbation growth rate. Shown in Fig. 2.8 are the 148-yr averaged S1s over all initial conditions for different calendar months and lead times (1, 3, 6 and 9 months). Note that in Fig. 2.8, S1s are from the SV analysis with perturbation of only SSTA. The amplitude of S1s with perturbations of both the SSTA and anomalous winds is almost the same as that of Fig. 2.8, therefore, not shown here. As can be seen in Fig. 2.8, large S1s often occur at their verification time (the end of the forecast) from August to October in corresponding predictions starting in the boreal spring or summer. For example, the maximum S1 occurs in September or October for 3, 6, and 9-month leads, corresponding to the starting month of June, March and February, respectively. This seasonal dependence in perturbation growth might explain why ENSO prediction skill often drops remarkably when prediction periods start from the boreal spring and pass through the boreal spring and summer, i.e., the ‘Spring Barrier’. Jin et al. (2008) recently performed 22-yr retrospective ENSO predictions using ten different coupled GCMs. Their results show that the

skill of forecasts that start in February or May drops faster than that of forecasts that start in August or November because predictions starting from February or May contain more events in the decaying phase of ENSO. Based on dynamics, the “Spring Barrier” is probably due to the fact that the Intertropical Convergence Zone (ITCZ) is closest to the equator during the spring, sustaining the unstable condition, and the ocean-atmosphere interaction is strong during the summer due to the relatively large vertical temperature gradient and ocean upwelling (e.g., Xue et al. 1997a). In addition, Fig. 2.8 shows that the magnitude of S1 increases with the lead time as expected.

It is interesting to explore whether S1 shows interannual or even longer time scale variability given the existence of decadal/interdecadal variations of ENSO prediction skill (e.g., Kirtman and Schopf 1998; Tang et al. 2008; Chen et al. 2004). A low-pass filter (2-yr) based on the Fourier transform (FFT) has been applied to the S1 of 6-month leads and the NINO3.4 SSTA index to address interannual and longer signals. The two filtered time series are shown in Fig. 2.9a and Fig. 2.9b. As can be seen, they have both visible interannual and longer time scale variability. The interannual and decadal/interdecadal variability of S1 and the NINO3.4 SSTA index can be further verified by the wavelet analysis shown in Fig. 2.9c and Fig. 2.9d. The local significant period varies between 2 and 20 years during the whole period from 1856-2003, including the ENSO interannual time scales (2-8 yr) and the decadal/interdecadal modulation of ENSO. On the interannual time scale, the significant time period of S1 tends to shift from a longer time scale to a shorter time scale. For example, the significant time period shifted from 8-yr to 3-yr between 1870 and 1900; this phenomenon reoccurred between 1900 and 1960. When comparing Fig. 2.9c with the wavelet power spectrum of the NINO3.4 index in Fig. 2.9d, similar shifting characteristics are found indicating that the changes in the significant periods of

the perturbation growth rate was associated with changes in the ENSO signal (spectrum power of NINO3.4) on the interannual time scale. It also indicates that ENSO tended to happen more frequently in recent decades and has a higher frequency of error occurrence. However, on the decadal/interdecadal time scales, Fig. 2.9c and Fig. 2.9d show that the power spectrums of S1 and NINO3.4 were not consistent in most of the time period except during the time period of 1900-1920 and around 1980. ENSO decadal/interdecadal signals were relatively weak between 1945 and 1975 while the spectrum power of the perturbation growth was significantly stronger over this period.

To examine relationships between NINO3.4 SSTA index and the S1, in particular their local relative phases, in time frequency space, the cross-wavelet analysis method (Grinsted et al. 2004) is applied for Nini3.4 SSTA and S1. The temporal variation of cross wavelet power spectrum is shown in Fig. 2.10, where the relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left, and NINO3.4 SSTA index leading S1 by 90° pointing straight down. As can be seen, both the phase synchrony and phase asynchrony between the two series can be observed at different time scales from decades to decades. For example, in-phase relationships are visible at the interannual time scales from 1880-1920 and 1940-1950 whereas the anti-phase relationships occurred at decadal/interdecadal time scales from 1900-1940 and 1960-1980. The anti-phase feature at decadal/interdecadal scales is in agreement with the ENSO predictability study in Tang et al. (2008) using multiple models, where they found that at decadal/interdecadal scales, strong ENSO events were related to small perturbation growth rates and vice versa. In addition, at interannual time sales, the significant periods seem gradually shifted to shorter scales from 1880-2000, which is probably due to the enhancement of ENSO variability in amplitude and frequency with time during the past 100

years. We will further discuss the relationships between ENSO signals and perturbation growth rate in part II of this work.

Many recent 20-30 yrs interval SV analyses concluded that: i) a small perturbation growth rate often occurs during an ENSO peak phase; and ii) the larger perturbation growth rate shows in the neutral and onset/breakdown stages of ENSO (Chen et al. 1997; Xue et al. 1997a; Tang et al. 2006; Zhou et al. 2008). Cai et al. (2003) obtained similar results when they analyzed the perturbation growth rate of the ZC model using a very long period breeding vector analysis. For comparison, we examined the above features of perturbation growth rate and ENSO phase over 148 yrs, resulting in a similar plot to Fig. 2.5 in Cai et al. (2003), as shown in Fig. 2.11. The ENSO events are binned into 18 categories between -2°C and 2.5°C with a 0.5°C interval based on the NINO3.4 SSTA index or NINO3 SSTA index (5°S - 5°N , 90°W - 150°W). The mean S1 of each category is shown as a function of the ENSO phase and the SSTA tendency. As shown in Fig. 2.11a, where 18 bars represent the 18 categories from the left to right, bins 1-9 have positive tendencies of SSTA and bins 10-18 have negative tendencies. In addition, bins 1-3 and 16-18 are at cold ENSO phase, bins 4-5 and 14-15 are at neutral phase, and bins 6-13 are at warm phase. The small perturbation growth rate occurs at the peak ENSO stage (peak El Niño and La Niña, bins 8-10 and 1, 18, respectively). While the large perturbation growth occurs prior to the decay phase of El Nino (bins 11-13) and during the transition period from a cold to a warm state (bins 3-5). These results are generally consistent with former SV studies (e.g., Chen et al. 1997; Xue et al. 1997a) and breeding vector results (e.g., Cai et al. 2003; Tang and Deng 2009) and further confirm the sensitivity of perturbation growth on ENSO phase. In the next section, we will identify and investigate the possible physical processes controlling the perturbation (error) growth in the ZC model.

2.4. Physical processes of perturbation growth in the ZC model

The evolutions of the model initial perturbations and ENSO signals are simultaneously controlled by internal dynamical and thermodynamical processes of the model such as the horizontal advection and vertical mixing. To explore underlying physical processes of the model perturbation growth, we decomposed the model SSTA variations into linear terms and nonlinear terms following the definition of An and Jin (2004), and performed several sensitivity experiments of SV analysis to investigate the contribution of individual term to the original total perturbation growth.

The governing equation of SSTA in the ZC model can be written as below

$$\frac{\partial T'}{\partial t} = -\vec{U} \cdot \nabla T' - U' \cdot \nabla (\bar{T} + T') - [M(\bar{W} + W') - M(\bar{W})] \frac{\partial \bar{T}}{\partial z} - M(\bar{W} + W') \frac{\partial T'}{\partial z} - \alpha T' \quad (2.9)$$

where $T'(\bar{T})$, $U'(\bar{U})$, and $W'(\bar{W})$ are anomalies (mean) of SST, surface layer currents, and vertical velocity, respectively, and α is a thermal damping coefficient. The first two terms on the right hand side of (9) are the horizontal advection terms. The third and fourth terms represent the effects of anomalous upwelling in the presence of the mean vertical temperature gradient $\frac{\partial \bar{T}}{\partial z}$, and the total upwelling in the presence of the anomalous vertical temperature gradient $\frac{\partial T'}{\partial z}$, respectively. The final term is a linear damping term, which can be interpreted as the change of SSTA due to the heat exchange between ocean and atmosphere. $M(x)$ is a step function: $M(x) = x$ if $x \geq 0$; $M(x) = 0$ if $x < 0$ which brings a cooling effect when there is upwelling and no effect otherwise since downward motion causes no change in SSTA.

If we consider the linear and nonlinear heating effects of horizontal advection and vertical advection (upwelling or downwelling), Eq. (9) can be expressed as:

$$\begin{aligned} \frac{\partial T'}{\partial t} = & -\bar{u} \cdot T'_x - \underline{u' T'_x} - \bar{u}' T'_x - \bar{v} T'_y - \underline{v' T'_y} - \bar{v}' T'_y - \underline{v' T'_y} \\ & - M(\bar{w}) T'_z - \{M(\bar{w} + w') - M(\bar{w})\} \bar{T}'_z \\ & - \{M(\bar{w} + w') - M(\bar{w})\} T'_z - \alpha T' \end{aligned} \quad (2.10)$$

where T , u , v , and w are SST, zonal, meridional, and vertical current velocities, respectively. The overbar and prime denote the climatological mean and anomaly, respectively. The underlined terms are nonlinear heating (NH) and the remaining terms are linear heating (LH), following the definition of An and Jin (2004). The linear and nonlinear heating terms can further be subdivided into the horizontal linear (HL), the horizontal nonlinear (HN), the vertical linear (WL), and the vertical nonlinear (WN), respectively. The linear dumping term is considered in the horizontal linear term (HL).

To identify the contribution of each individual heating term to the original perturbation growth, we performed SV analysis for each linear and nonlinear term over 148 years respectively. Note that the nonlinear heating terms have been linearized in the TLM, the nonlinear perturbation/perturbation growth mentioned hereafter are actually the linearized nonlinear perturbation contributions. The SV analysis of each term is similar to the original analysis described in section 2.2 except that the perturbation growth of SST (i.e., X_f in Eq. 2.7) was replaced by the perturbation growth of an individual heating term obtained from the TLM. This is confirmed by the results obtained using tangent linear model, where the original SV1 is used as initial condition X_i for integration of tangent linear model. It was found that SV1s of these heating terms are similar to the original SV1 as shown in Fig. 2.1a. This is because the

solution of maximizing total perturbation growth rate A in Eq. (2.4) is equivalent to the solution of maximizing growth rate of each individual term. The final patterns of these terms are subject to their physical processes, representing the perturbation contribution from each heating term. The final pattern of each term from SV analysis is actually equivalent to the response of corresponding term (X_j) to the original SV1 (X_i) by (Eq. 2.7). The 148-yr averaged final patterns for the linear and nonlinear terms are given in the left and the right panel of Fig. 2.12, respectively. The final pattern of the total horizontal linear heating (Fig. 2.12e) is very similar to the original final pattern in Fig. 2.1b. There are two positive perturbation growth regions in the tropical equatorial Pacific, located in the central Pacific Ocean and the eastern Pacific, respectively. The former center in the central Pacific near 150°W , where the strong atmosphere-ocean interactions and large instability conditions often occur, is formed as a result of the horizontal linear perturbation growth (see Fig. 2.12a). The perturbation growth in the eastern Pacific is clearly related to the vertical linear term (Fig. 2.12c), indicating that the optimal growth in the eastern Pacific Ocean is mainly due to the vertical linear upwelling/downwelling term. This vertical linear optimal growth is probably due to an inaccurate parameterization of the vertical mixing process. The high spatial similarity of the final patterns of the total linear and horizontal linear optimal growth suggests the linear heating perturbation growth dominates the total model perturbation growth. However, if we ignore the perturbation contribution of the nonlinear process, the perturbation growth in the central Pacific would be much stronger than the perturbation growth in the original final pattern. This large perturbation growth in the linear process implies that there must be some offset effects (negative optimal growth) in the total nonlinear heating processes that reduce the large linear perturbation growth. We can see this reduction in the total nonlinear perturbation in Fig. 2.12f: there is a negative perturbation

growth center in the central Pacific near the dateline region where it can partly offset the positive perturbation growth in the total linear perturbation. Therefore, both the linear perturbation growth and nonlinear perturbation growth are important in the central Pacific. The total nonlinear error can be further decomposed by the horizontal nonlinear term and the vertical nonlinear term shown in Fig. 2.12b and Fig. 2.12d. There is a large negative perturbation growth region in the central Pacific in the horizontal nonlinear term, which is similar to the total nonlinear perturbation growth pattern shown in Fig. 2.12f; meanwhile, a relatively weak positive perturbation growth is shown in the vertical nonlinear term in the central Pacific (Fig. 2.12d). Therefore, the total nonlinear negative error is mainly the result of the horizontal nonlinear term. Comparing Fig. 2.12e with Fig. 2.12f shows the perturbation growth contribution of the total linear heating is 3-4 times larger than the contribution of the nonlinear heating (note that the perturbation growth rates were included in FPs). Therefore, total model errors are mainly caused by the linear advection heating process, but the linear process can be partially offset by the nonlinear process which has a negative error contribution, especially in the central Pacific.

To compare the error contributions of individual linear/nonlinear heating terms, the seasonal variations of these perturbation growth rates are given as a function of the forecast verification time (Fig. 2.13). As expected, the horizontal linear heating term (HL) makes the largest contribution to the original growth rate $S1$, and shows a consistent seasonal variation with original perturbation growth $S1$. The vertical linear heating (WL) and horizontal nonlinear heating (HN) have comparable error contributions, but they are much smaller than horizontal linear perturbation growth (HL). Comparing the vertical nonlinear perturbation growth (WN) with the other three terms shows that the vertical nonlinear perturbation growth is the smallest

contributor with very weak seasonal variations. To visualize the linear and nonlinear error contributions more clearly, seasonal variations of the perturbation growth rates are given in Fig. 2.13b. The perturbation growth of the total linear term is about twice as large as the total nonlinear perturbation growth, which confirms again that the original total perturbation growth S_1 is mainly determined by the linear process, and the nonlinear process contributes to a smaller and negative perturbation growth. An offsetting effect between the linear and nonlinear terms explains why the horizontal perturbation growth rate HL is larger than the original perturbation growth rate S_1 .

To further understand underlying mechanisms of linear and nonlinear perturbation growth, we performed several EOF analyses for individual linear/nonlinear heating terms to look for dominating physical processes that control the variation of total heating, and investigate the relationship between the perturbation growth rate and the corresponding heating term. These individual heating terms were obtained from the integration of the original model again over the period of 1856-2003. For the total linear heating process, the first EOF mode, accounting for 73.1 % of total variance, shows an ENSO-like pattern (Fig. 2.14c). Comparing this EOF mode of total linear heating with the horizontal linear heating (Fig. 2.14a) and vertical linear heating (Fig. 2.14b), reveals that the warming in the equatorial central and eastern Pacific is from the contribution of anomalous horizontal linear heating, and the warming along the coastal zone is mainly due to vertical linear heating. From the corresponding principal components (PCs) shown in Figs. 2.14d-f, linear heating is more likely to cause warming as indicated by dominant positive values in the PCs. Figs. 2.15a-c are the first EOF modes of the horizontal, vertical, and total nonlinear heating terms respectively. Their corresponding PCs are given in Figs. 2.15d-f. EOF analyses show a cooling and warming pattern for horizontal and vertical nonlinear terms

in Fig. 2.15a and Fig. 2.15b, respectively. Considering that all PCs are positive and that the horizontal nonlinear PC has a larger amplitude than the vertical nonlinear term, the total nonlinear heating NH can be explained by the horizontal cooling effect as shown in Fig. 2.15c. Xue et al. (1997a) obtained similar results from an older version of the ZC model with a shorter time period, and concluded that the horizontal nonlinear advection is mostly a cooling effect and the vertical advection is mostly a warming effect, namely that, the vertical nonlinear advection always strengthens warm SST anomalies but diminishes cold SST anomalies in the eastern Pacific.

These nonlinear vertical warming and horizontal cooling effects can be further explained mathematically by Eq. (10) together with the final patterns of SSTA and the wind field in Fig. 2.2. For example, during an El Niño event, the easterly trade wind is weakened and a westerly current anomaly ($u' > 0$) occurs in the central Pacific. Meanwhile, the horizontal SSTA warming increases from the west to the east showing a positive zonal SSTA gradient ($T'_x > 0$). Thus, the horizontal nonlinear advection ($-u'T'_x < 0$) contributes a cooling effect in the central Pacific. This horizontal nonlinear cooling/dumping effect in the ZC model is in agreement with An and Jin's report (2004) that during the developing phase of El Niño, both the anomalous zonal temperature gradient and the anomalous zonal current in the surface layer are positive, which leads to a negative nonlinear zonal advection. On the other hand, a weakening of upwelling ($w' < 0$) and a stronger warming at the sea surface than in lower layers ($T'_z > 0$) are found in the ZC model (opposite w' and T'_z in observations in An and Jin (2004)). Therefore, the nonlinear vertical mixing ($-w'T'_z > 0$) contributes to a warming effect in the central Pacific, which can partly offset the horizontal nonlinear cooling. For La Niña events, in the central

Pacific $u' < 0$, $T'_x < 0$ along with a strengthened upwelling, $w' > 0$, and a cooling sea surface $T'_z < 0$. Therefore, the horizontal nonlinear cooling and vertical nonlinear warming are valid.

However, comparing the horizontal and vertical nonlinear heating terms in the ZC model with that from the observations in An and Jin (2004) shows some physical deficiencies of the ZC model: i) The model's vertical nonlinear term does not show a great enough warming effect to offset the horizontal nonlinear cooling contribution, therefore, the net nonlinear heating is a cooling effect, whereas, in the observations of An and Jin (2004) the vertical nonlinear warming dominates the net nonlinear heating; ii) The vertical nonlinear warming in the model is located in the central Pacific, while the warming dominated in the eastern Pacific near the cold tongue region in An and Jin (2004); iii) Although there is an out-of-phase relationship between the upwelling (w') and the vertical temperature gradient ($T'_z = SSTA - T_{sub}$) through the ENSO cycle in the model, both signs of the vertical motion and the temperature gradient in the model are opposite to observations. The model has a weakening of upwelling ($w' < 0$) in El Niño events and the subsurface warming is smaller than the surface warming $T'_z > 0$. However, there is a strong warm water upwelling occurring in the eastern Pacific in the observations, especially for those strong El Niño events after 1980. An et al. (2005) compared nonlinear heating terms in 10 coupled models and found only one model gave the correct simulation. Most models did not represent both the location and strength or even the sign of the nonlinear vertical warming. This model bias in the internal model dynamics and physical processes certainly will cause perturbation growth but that is beyond the scope of this paper.

The spatial patterns of linear and nonlinear heating terms revealed by EOF analysis are very similar to their corresponding final patterns of perturbation growth. Furthermore, over 148 yrs,

significant positive correlations have been found between each PC and its corresponding singular value. Strong linear heating is associated with a faster positive perturbation growth while strong nonlinear heating leads to a faster negative perturbation growth. For example, the correlation coefficients between PCs of linear terms HL/WL/total linear and their corresponding singular values are 0.56/0.46/0.44 over 148 yrs, which are all statistically significant at the 99% confidence level. The correlation coefficients between PCs of nonlinear terms HN/WN/total nonlinear and their singular values are 0.54/0.66/0.46, respectively. Very high spatial similarity and temporal correlations between each perturbation growth rate and the corresponding heating term suggest that the linear perturbation growth (L1) and nonlinear perturbation growth rate (N1) are highly related to the linear/nonlinear physical processes themselves. Comparing Fig. 2.15 with Fig. 2.14 reveals that the total linear heating makes a larger contribution to the total heating, leading to the finding that linear processes contribute more to the total perturbation growth than nonlinear processes as found in Fig. 2.13.

2.5. Conclusion

It is important to identify a statistically robust SV analysis of ENSO prediction models. The relationship between singular value and ENSO predictability has not been sufficiently addressed in previous studies of ENSO predictability due to a lack of long term retrospective prediction and corresponding SV analysis. In this work, a tangent linear model is constructed for the latest ZC model version LDEO5 to study perturbation growth and ENSO predictability for the past 148 years from 1856-2003. It provides a substantial account of the error growth rate and spatial patterns in LDEO5 from seasonal to interdecadal time scales.

From the 148-yr singular vector analyses by our new constructed physical-based TLM, the

long-term averaged first singular vector SV1 is a west-east dipole spanning the equatorial Pacific with centers located in the east and the central Pacific Ocean. Comparing the SV1 of LDOE5 with that of the previous SV studies (i.e., Chen et al. 1997; Xue et al., 1997a), we find that the north-south dipole in the older ZC model version in the eastern Pacific is missing, which might be due to improvements in the ZC model (i.e., model dependent). A spatial correlation between the monthly SV1s and the 148-yr averaged SV1 agrees with previously published results showing that SV1 is less sensitive to model initial conditions while there is a strong sensitivity of singular values to initial conditions. The faster model perturbation growth during spring/summer is probably caused by the stronger atmosphere-ocean interaction. Besides the seasonal variations, the leading singular value, S1, has significant periods ranging 2-20 years as seen in the wavelet analysis. On the interannual time scales, the significant time scales of S1 and the ENSO signal occasionally shifted from longer periods to shorter periods during the 148 years.

The relative contribution of linear and nonlinear heating to S1 has not so far been addressed well. In this study, we also conducted SV analysis for each individual heating term in the SST governing equation. SV analyses on the individual linear and linearized nonlinear terms reveal that the model optimal perturbation growth is mainly from linear heating terms. The total linear optimal perturbation growth is twice as large as the total nonlinear term. The final optimal perturbation growth pattern of an individual heating term has a similar spatial pattern as the EOF pattern of the heating term. In addition, significant correlations have been found between the perturbation growth rate of each term and corresponding PC-1 of the EOF analysis for the individual heating terms. Therefore, the singular value of each heating term depends significantly on the heating term itself. The perturbation growth in the central equatorial Pacific, where strong

atmosphere-ocean interaction occurs, is dominated by a positive perturbation growth from the horizontal linear term. The perturbation growth in the eastern Pacific is dominated by vertical linear mixing, which is probably related to inaccurate parameterization of the mixing process.

A robust and stable optimal error growth pattern, SVI, and the optimal error growth rate, S1, over 148 years will be useful indicators of potential predictability. Further discussion of the relationship between potential predictability that does not use observations, and the model prediction skills indicated by correlation and root mean square error (RMSE) from a comparison with observations will be presented in Part II of this study. The relationship will offer a practical means of estimating the confidence level of ENSO prediction using the dynamical model. In addition, the SVIs obtained in the present study provide an optimal tool to construct ensemble predictions, i.e., repeating a prediction many times by perturbing the initial conditions of a forecast model with SVs and random noise each time. Through statistical predictability theory and ensemble prediction of the past 148 years, the shape of the forecast probability density function (PDF) that describes the prediction uncertainty can be estimated, and the nature of ENSO predictability explored in the chapters 3-5 of this study.

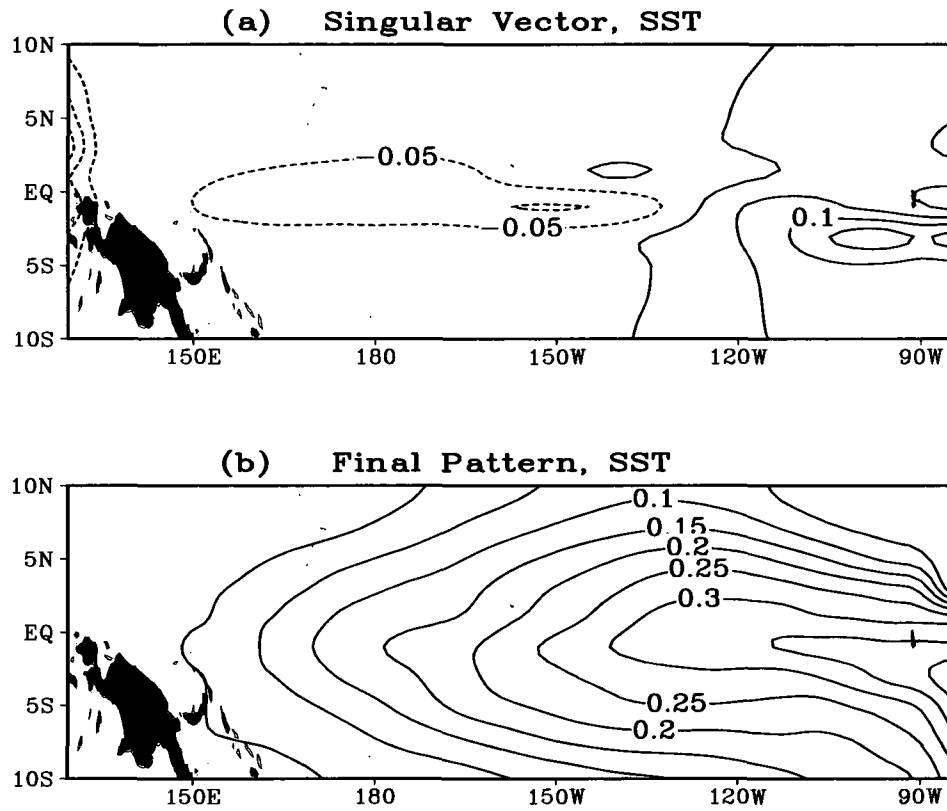


Fig. 2.1 The first singular vector (SV) and the first final pattern of SSTA averaged in the 148 years, a) the first singular vector (SV) of SSTA; b) the first final pattern of SSTA. (SV1 & FP explain 32 % of the variance of R in the SVD analysis using Eq. 2.5) (Unit: $^{\circ}\text{C}$)

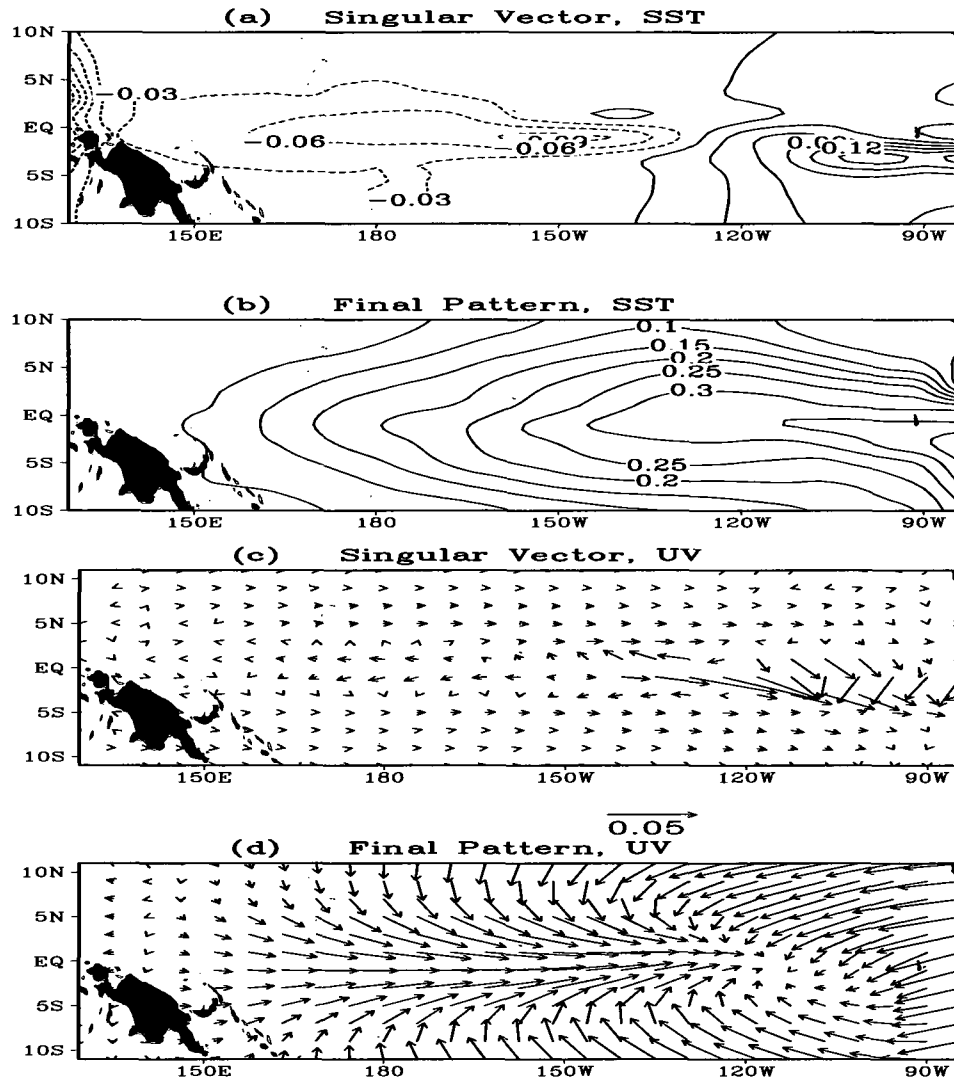


Fig. 2.2 Same as Fig.2.1 but perturbing both SSTA ($^{\circ}\text{C}$) and wind field (m/s). a) The first singular vector of SSTA; b) the first final pattern of SSTA; c) the first singular vector of the wind field; d) the first final pattern of the wind field.

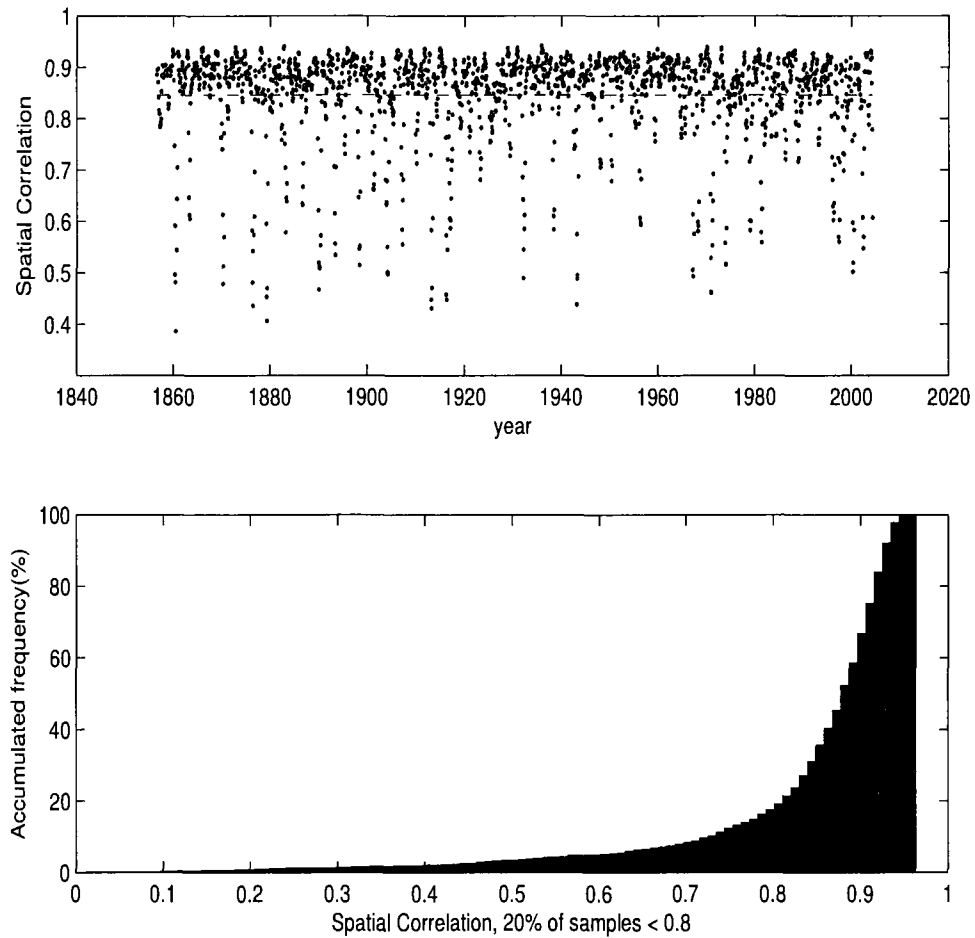


Fig. 2.3 a) Spatial correlation coefficients between each first SV over 148-yr and the averaged SV-1, the mean correlation coefficient is 0.85 (dash line); b) Accumulated frequency against the spatial correlation coefficient. It indicates the fraction of SVs that is smaller than the spatial correlation coefficient. For example, 20% of SV samples have the spatial correlation less than the spatial correlation of 0.8.

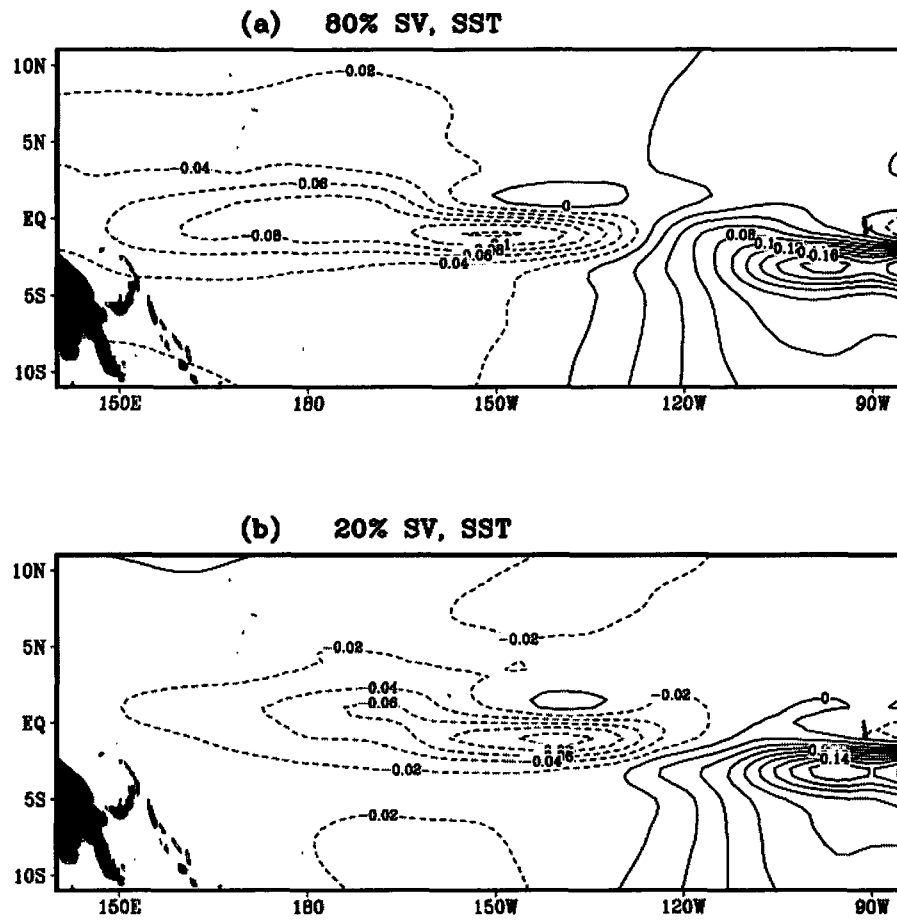


Fig. 2.4 a) Composite SV1 of SSTA ($^{\circ}\text{C}$) for (a) high spatial similarity cases (80% of total SV1) and (b) low spatial similarity (20% of total SV1).

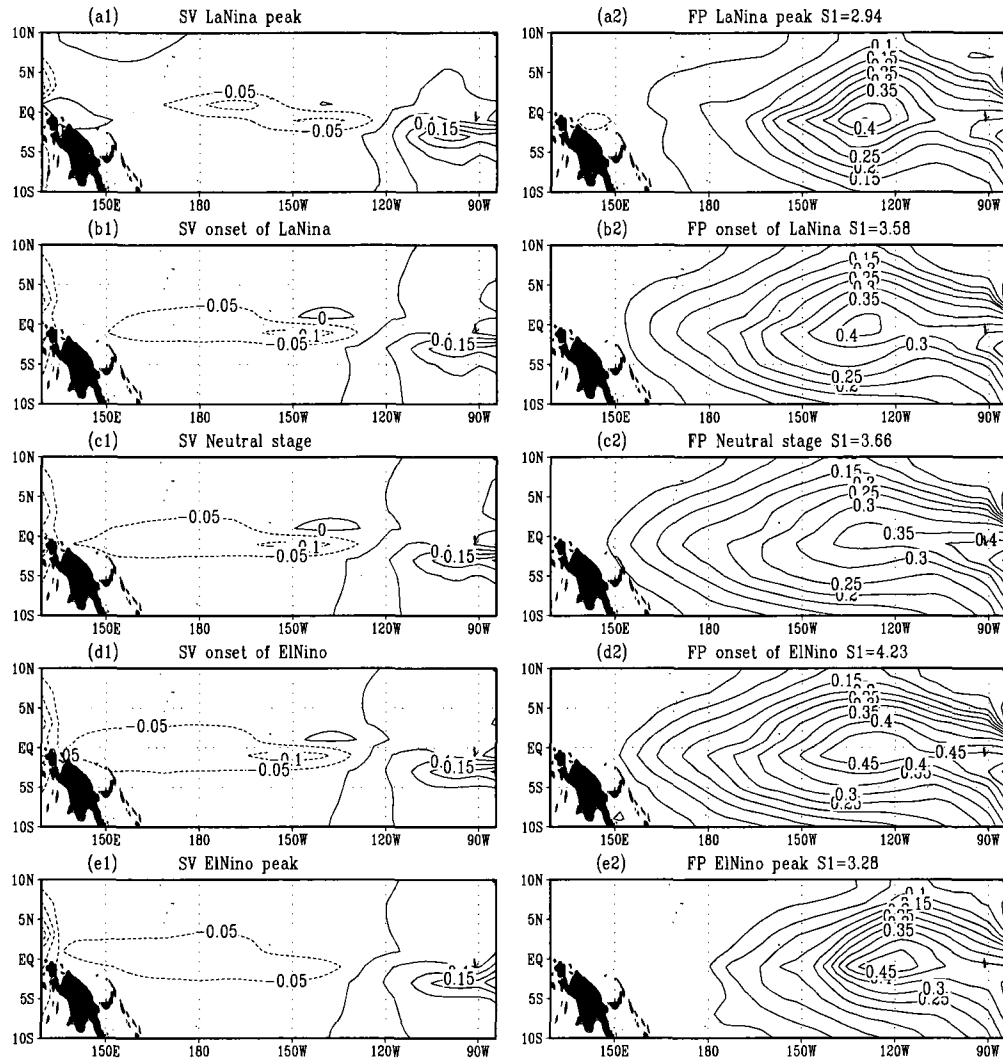


Fig. 2.5 the first SV of SSTA (°C) starting from the phase of a1) peak La Niña; b1) onset of La Niña; c1) Neutral; d1) onset of El Niño; e1) Peak El Niño. The corresponding final pattern after 6 months is shown in the right panel a2-e2. The averaged perturbation growth rate $S1$ of each stage is marked in the title captions.

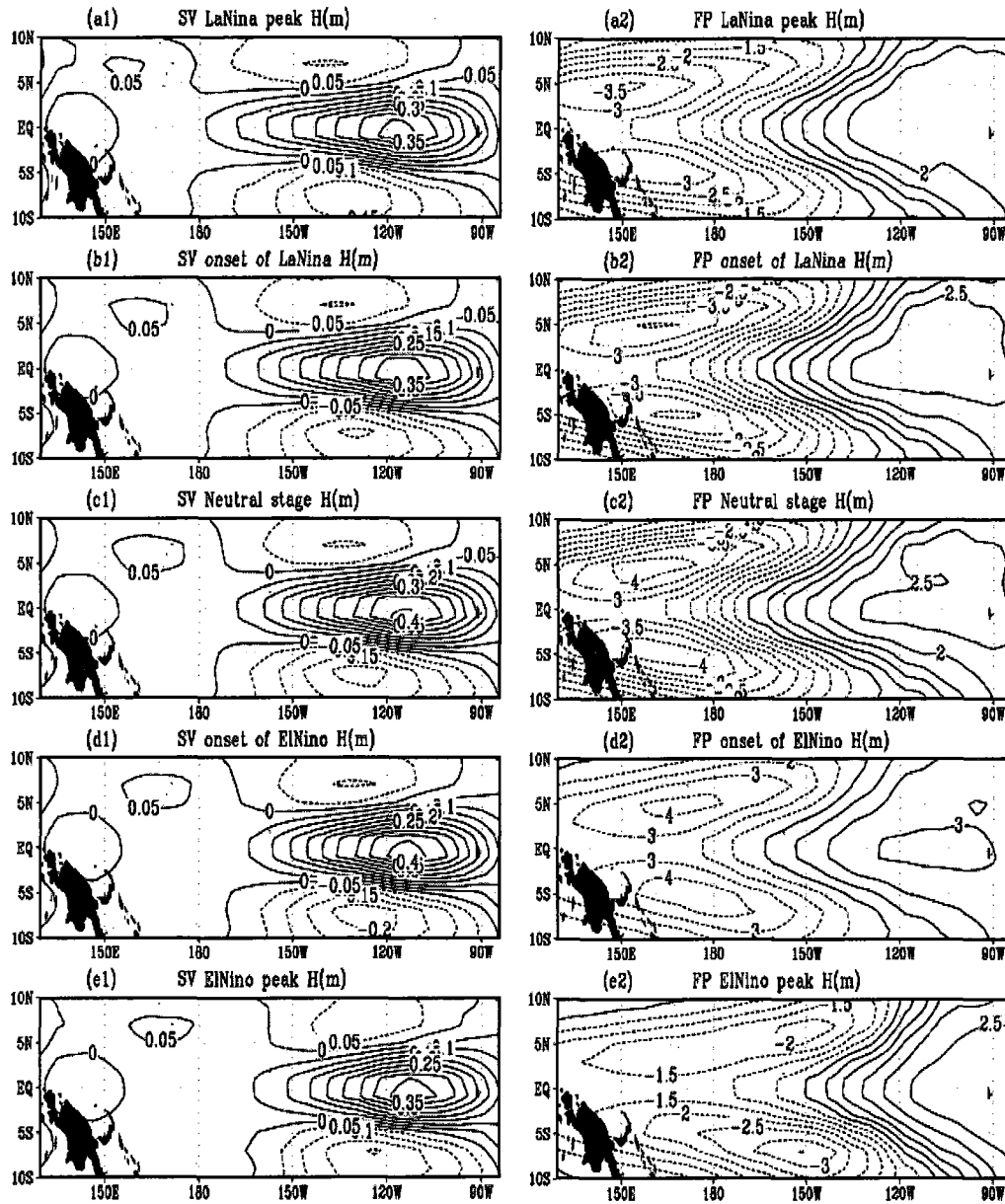


Fig. 2.6 Same as Fig. 2.5 but for thermocline depth anomaly (H) (unit: m).

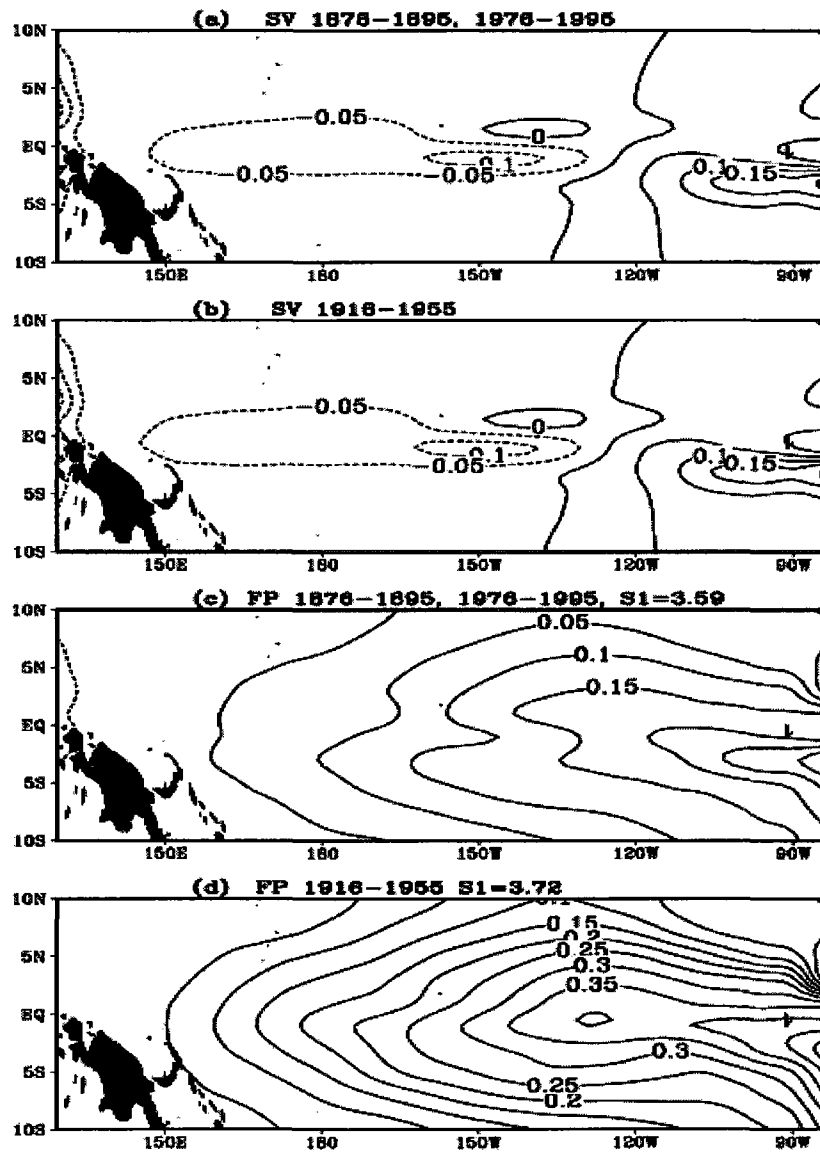


Fig. 2.7 a) the SV and b) the final pattern (FP) averaged in the higher skill period (1876-1895 and 1976-1995) and c) SV and d) the final pattern in the lower skill period (1916-1955).

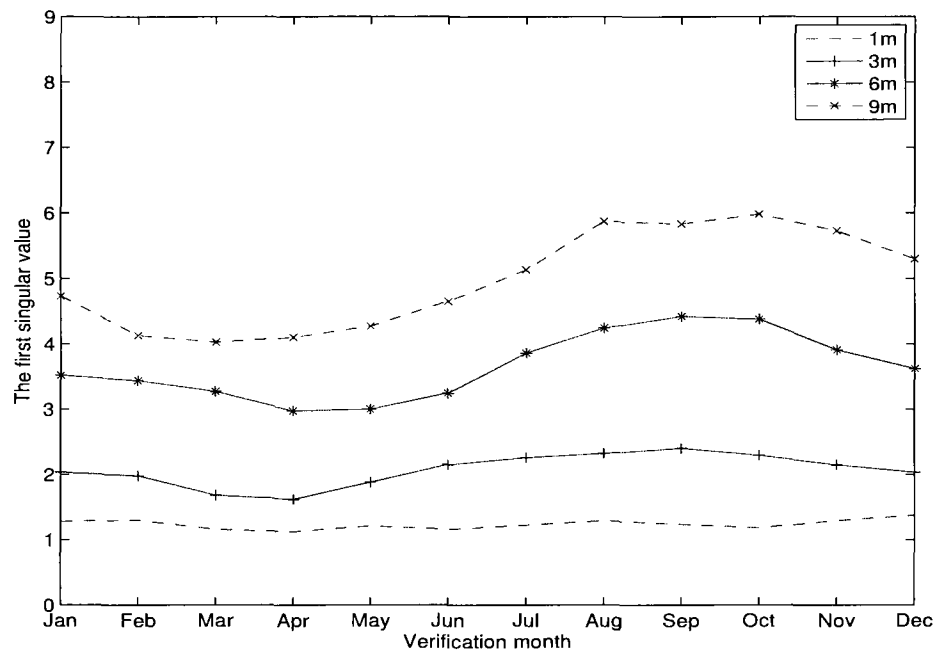


Fig. 2.8 Seasonal variations of the first singular values S1 against the prediction target time at different lead time, 1, 3, 6 and 9 months (from bottom to top respectively) averaged over 148 years.

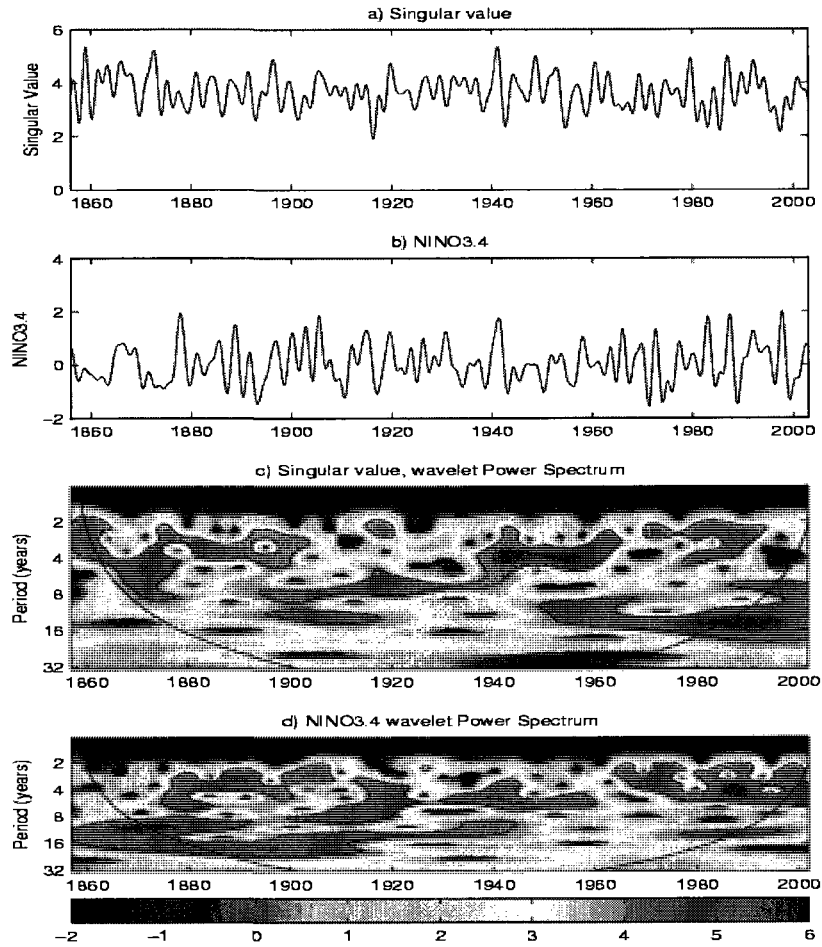


Fig. 2.9 Time series of the low-pass filtered (>24 months) a) the first singular value (S1) and b) the NINO3.4 index used for the wavelet analysis. Wavelet power spectrum of c) S1 and d) NINO3.4 using the Morlet wavelet. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The solid smooth curves in the left and right corners indicate the edge effects become important.

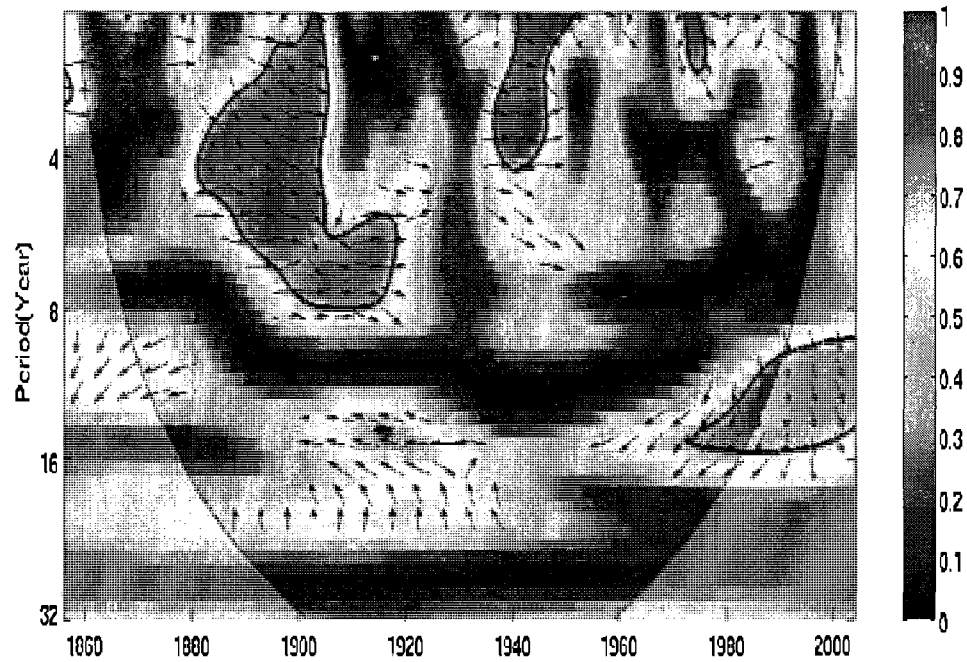


Fig. 2.10 The cross-wavelet analysis for NINO3.4 SSTA index and the singular value S1. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left, and NINO3.4 SSTA index leading S1 by 90° pointing straight down.

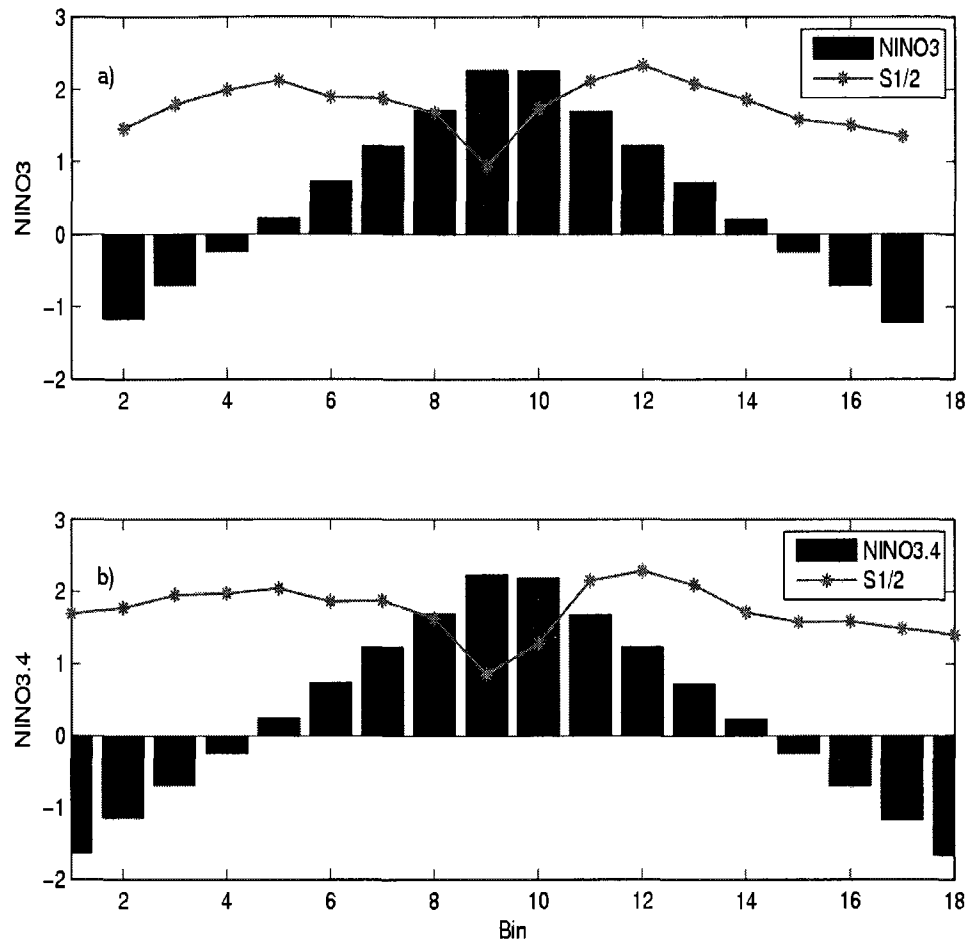


Fig. 2.11 Mean S1 (in solid star curve) as a function of the background ENSO phase. a) The bar curve is the NINO3.4 index of the composite background ENSO cycle (S1 is divided by 2); b) same as a) but using NINO3 index.

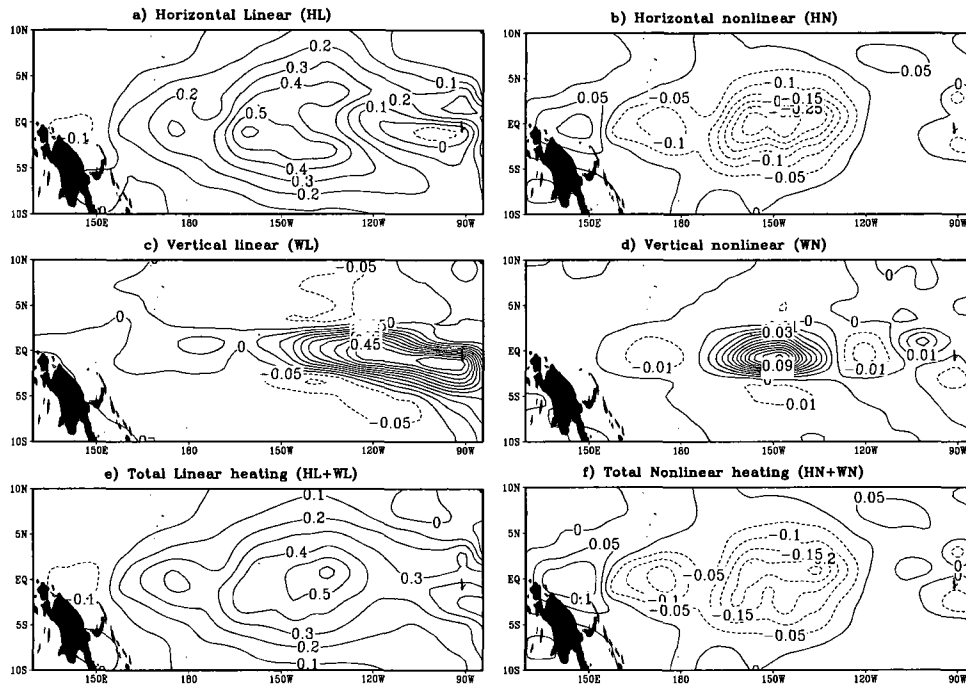


Fig. 2.12 Final patterns for linear and nonlinear heating terms averaged in 148 years: a) the final pattern of horizontal linear heating (HL); b) the final pattern of horizontal nonlinear heating; c) the final pattern of vertical linear heating (WL); d) the final pattern of vertical nonlinear heating (WN); e) the final pattern of total linear heating (HL+WL); f) the final pattern of total nonlinear heating (HN+WN). (Unit: $^{\circ}\text{C}$)

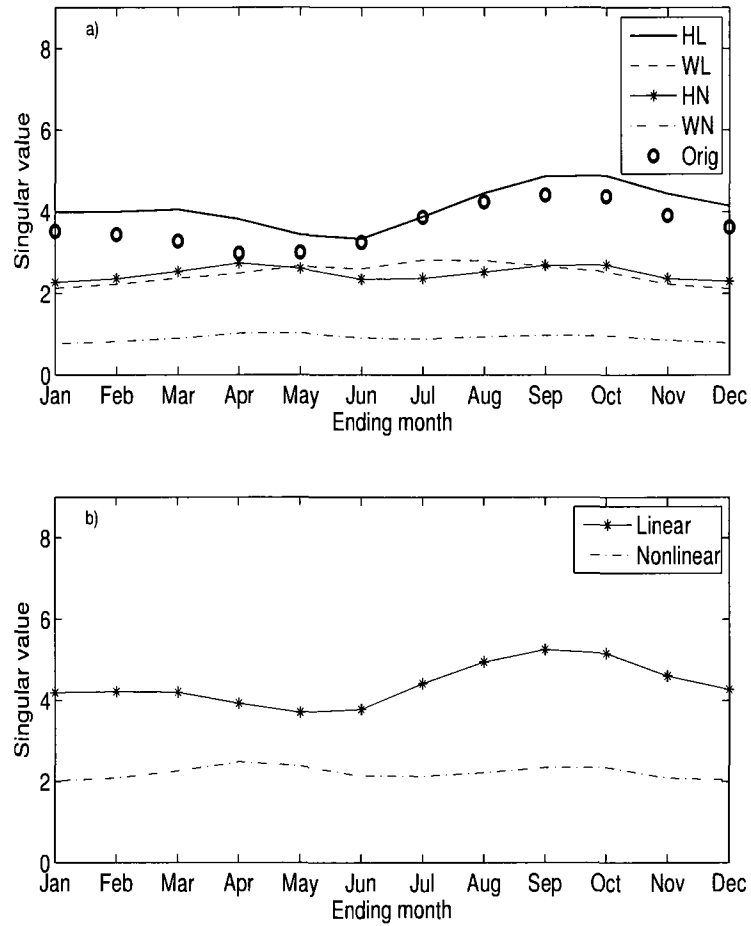


Fig. 2.13 The seasonal variation of singular values for linear and nonlinear heating terms a) horizontal linear heating (HL, solid line), vertical linear heating (WL, dash line), horizontal nonlinear heating (HN, solid star line), and vertical nonlinear heating (WN, dash dot line). The original singular values (open circles); b) singular values for the linear heating (solid star line), nonlinear heating (dash dot line).

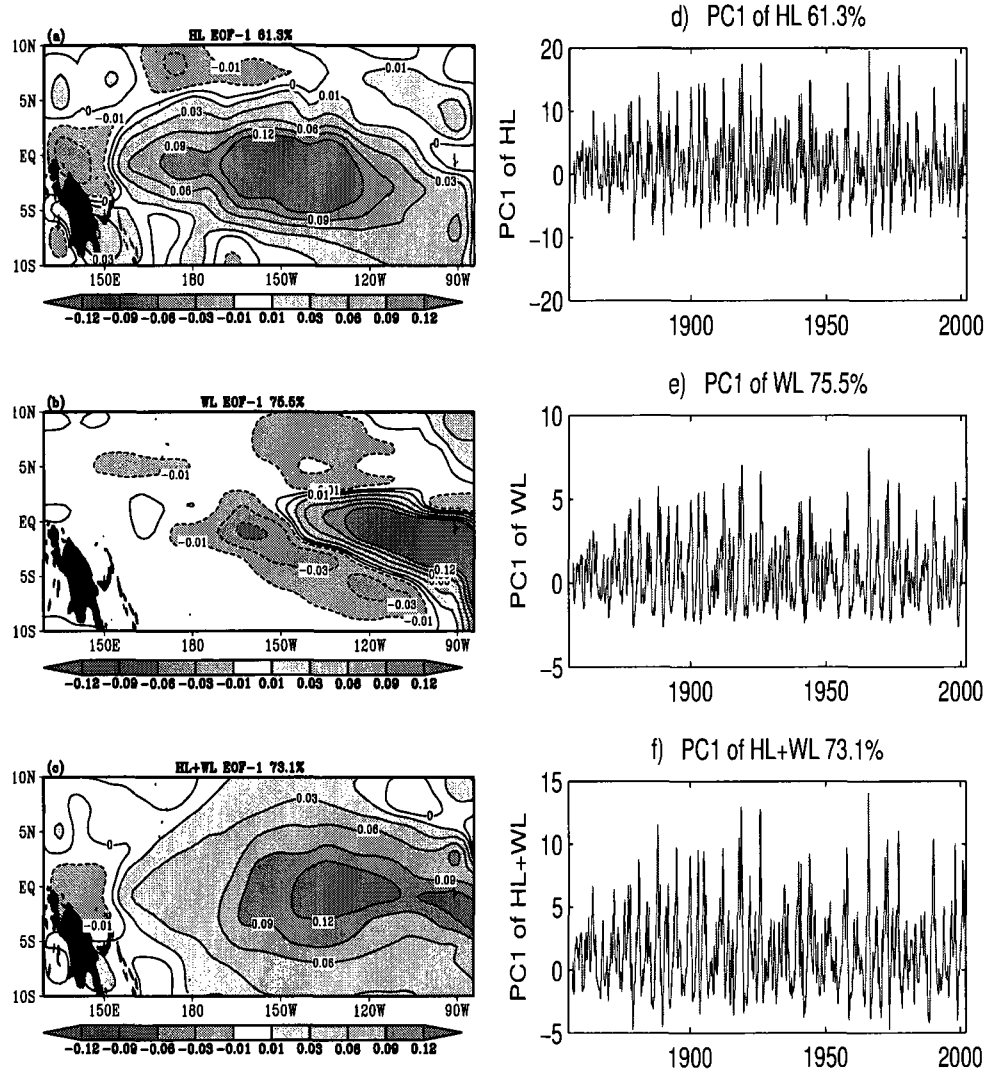


Fig. 2.14 EOF analyses for linear heating terms in Eq. (10). From top to bottom, Figures in the left panel are the EOF-1 spatial patterns of a) horizontal linear (HL), b) vertical linear (WL) and c) total linear heating (HL+WL). Their corresponding PCs are in the right panel. (Unit: °C/month)

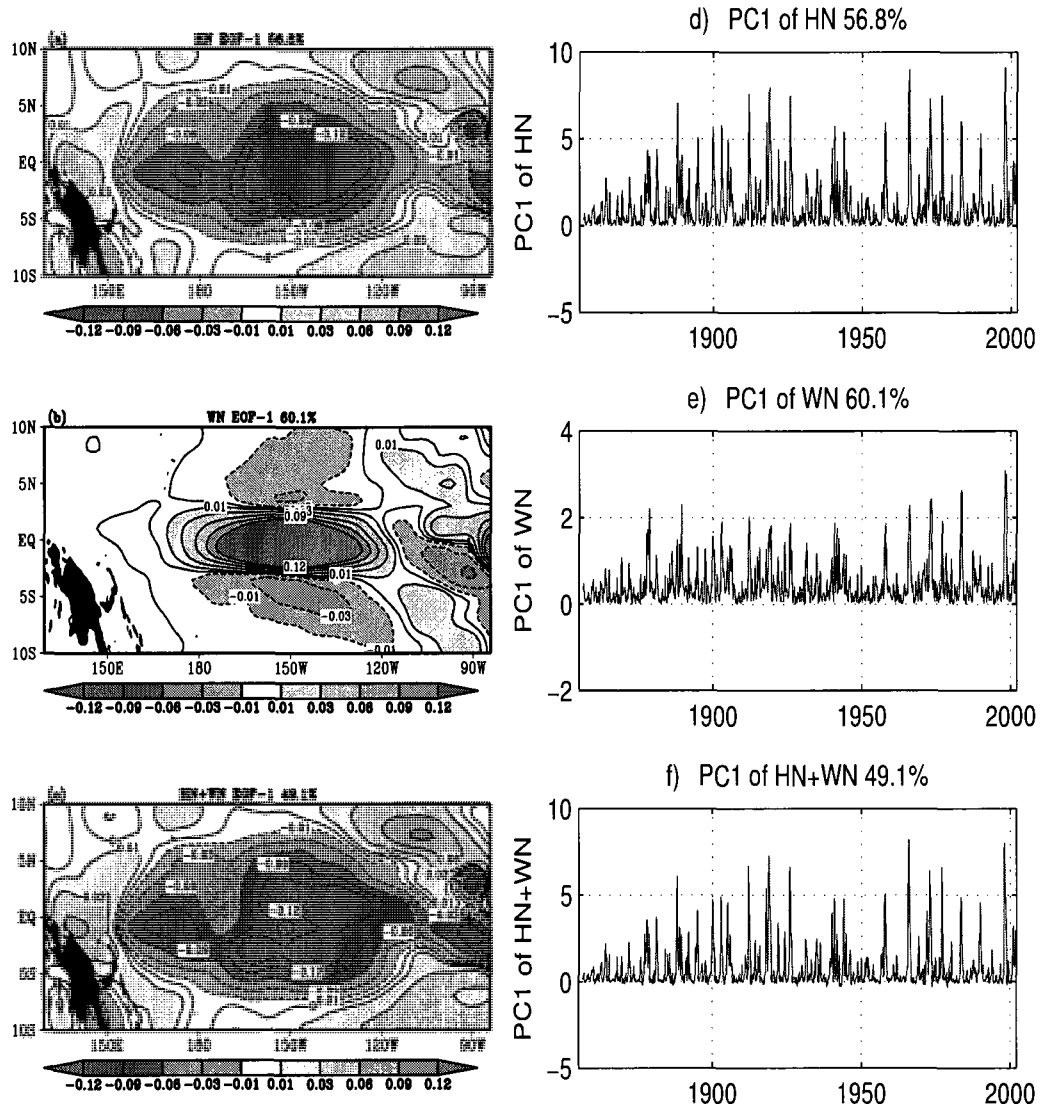


Fig. 2.15 Same as Fig. 2.14 but EOF analyses results for nonlinear heating terms. Figures from top to bottom are a) horizontal nonlinear (HN), b) vertical nonlinear (WN) and c) total nonlinear heating (HN+WN).

Chapter 3: Further Analysis of Singular Vector and ENSO Predictability in the Lamont Model --- Part II: Singular value and predictability

Cheng Y, Tang Y, Jackson P, Chen D, Zhou X, Deng Z (2009) Further analysis of singular vector and ENSO predictability in the Lamont model—Part II: Singular Value and predictability. Climate Dynamics, DOI: 10.1007/s00382-009-0728-z.

Published version is available at:

<http://www.springerlink.com/content/k5g37856j3t55r70/>

3.1. Introduction

ENSO predictability displays multiple time scales in numerical models, including the seasonal, interannual, and decadal/interdecadal time scales. On the seasonal time scale, ENSO forecast skills in many models decline significantly in the boreal spring with apparent skill recovery in subsequent seasons, showing the “spring barrier” phenomenon (e.g., Jin et al., 2008). On the interannual time scales (2-7 yrs), ENSO prediction skills are associated with ENSO phase and ENSO intensity, namely, strong ENSO events have high prediction skills, while the neutral ENSO states have poor prediction skills (e.g., Tang et al. 2005; 2008a); The growth phases of both the warm and cold events are better predicted than the corresponding decaying phases in many coupled ENSO forecast models (e.g., Jin et al. 2008). These features of ENSO predictability also occur in the Zebiak-Cane model (Zebiak and Cane 1987; Chen et al, 2004; hereafter ZC); for example, the warm and cold events are equally predictable while near normal conditions are harder to predict (Chen and Cane 2008). On the decadal/interdecadal time scales, ENSO predictability has apparent decadal/interdecadal variations (e.g. Wang 1995; Kirtman and Schopf 1998; Latif et al. 1998; Chen et al. 2004; Tang et al. 2008a). Tang et al. (2008a) explored ENSO predictability using three models and long term retrospective predictions. Consistent results and conclusions were found in the three models with different complexity, namely, higher prediction skills for the late 19th century and late 20th century, and lower skills for the period of 1916-1955. These consistent relationships found in the three models offer valuable insight to some important issues of ENSO predictability on the longer time scales.

Typically, there are two hypotheses responsible for the loss of predictability with forecast

lead time. The first argues that the loss of predictability is due to the chaotic behavior of the nonlinear dynamics of the coupled system (e.g., Jin et al. 1994; Chen et al. 2004), whereas the second attributes it to the stochastic nature of the coupled system characterized by weather noise and other high-frequency variations, such as westerly wind bursts and the Madden–Julian oscillation (e.g., Kirtman and Schopf 1998; Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Moore and Kleeman 1999; Vecchi and Harrison 2003; Moore et al. 2006; Gebbie et al. 2007; Jin et al. 2007). It is still not clear to date which regime plays the dominant role in controlling the variation of ENSO predictability.

Singular vector analysis (SV) is a powerful tool to study predictability because the optimal perturbation growth suggests the intrinsic limits of prediction skill. The SV has been widely used to study the loss of ENSO predictability due to initial error/perturbation growth (i.e., Lorenz 1965; Chen et al. 1997; Xue et al. 1997a, b; Fan et al. 2000; Tang et al. 2006; Zhou et al. 2008). These SV analyses showed that the perturbation growth rate (i.e. singular value) is sensitive to the seasonal cycle, ENSO phase, and ENSO signals. However, all of the above studies focused on a period of only 20-40 years, with a rather limited number of ENSO cycles, basically precluding statistically robust conclusions. In theory, an inverse relationship could be expected between the leading growth rate and the ENSO predictability. Due to a lack of long term retrospective prediction and corresponding SV analysis, however, the relationship between the singular value and ENSO predictability has not been sufficiently addressed, and especially has not been validated by actual prediction skill measures in previous SV studies. Chen et al. (2004) performed a retrospective forecast experiment spanning the past 148 years, using only reconstructed SST data for model initialization. At a 6-month lead, the model was able to predict most of the warm and cold events occurred during this long period, especially for the

relatively large ENSO events. Using the long-term reconstructed SST data and the ZC model LDEO5 version, we recently completed a long-term SV analysis and corresponding retrospective ENSO prediction for the period from 1856-2003. In part I of this work (Chapter 2), we constructed a fully physically-based tangent linear model (TLM) for the ZC model, explored the variations of singular vectors and singular values in the time scales from seasons to decades, and examined the control factors responsible for SV variations over the 148 years. A robust and stable optimal perturbation growth pattern and the optimal perturbation growth rate for the 148 years were obtained in part I, which could be useful indicators of predictability. To extend this work, the present study focuses on exploring the relationships between the optimal perturbation growth rate, a potential measure of predictability which does not make use of observations, and ENSO actual prediction skills that do make use of observation, for the 148 years at multiple time scales ranging from the interannual time scale to decadal/interdecadal time scale. The identified relationship has a theoretical contribution to predictability study using SV, and a practical significance in estimating the confidence that we can place in future predictions using the same ENSO forecast model.

In section 2.2.1, we present a brief introduction to the LDEO5 model and the metrics used to measure actual ENSO prediction skill. The relationships between these prediction skill metrics and the perturbation growth rates are discussed in section 3.3 and section 3.4. The relationships between the actual prediction skill, perturbation growth rate, and ENSO signals are analyzed in section 3.5, followed by a conclusion and discussion in section 3.6.

3.2. Methods

A fully physically-based tangent linear model (TLM) was constructed for the LDOE5

model and singular vector analysis performed for the 148-year period from 1856 to 2003, as shown in Chapter 2. From the long-term SV analyses, the leading singular value (S1) that represents the optimal perturbation growth rate of forecast SSTA, the linear component of S1 (denoted by L1), and the nonlinear component of S1 (denoted by N1) for the 148 years have been obtained. We will use these perturbation growth rates as the potential predictability measures to investigate their relationship with several actual prediction skill measures for the LDEO5 model. The actual prediction skill metrics are discussed in section 3.2.

3.2.1 Metrics of Actual Prediction Skill

Traditionally, the actual prediction skill of ENSO is measured by anomaly correlation coefficient (R) and the mean square error (MSE) between predicted the Nino3.4 SSTA index (averaged over 5°N to 5°S, from 170°W to 120°W) against the observed counterpart.

$$R(t) = \frac{\sum_{i=1}^N [T_i^p(t) - \mu^p] [T_i^o(t) - \mu^o]}{\sqrt{\sum_{i=1}^N [T_i^p(t) - \mu^p]^2} \sqrt{\sum_{i=1}^N [T_i^o(t) - \mu^o]^2}} \quad (3.1)$$

$$MSE(t) = \frac{1}{N-1} \sum_{i=1}^N (T_i^p(t) - T_i^o(t))^2 \quad (3.2)$$

where T is the index of NINO3.4 SSTA, t is the lead time of the prediction from 1 to 12 months, T^p is the predicted NINO3.4 SSTA, and T^o is corresponding observed counterpart, subscript i the initial time of prediction ($i = 1, \dots, N$); μ^p is the mean of the forecasts, μ^o is the mean of observations. N is the number of samples used over 148 years in this study, a total of 148×12

($N=1776$) forecasts, initialized from January 1856 to December 2003, were run starting at one month interval (1 January, 1 February ...1 December), and continued for 12 months for the ZC model. SST assimilation was used to initialize the forecasts as discussed in Chen et al. (2004). The seasonal cycle has always been removed from forecasts and observations prior to measuring prediction skill. To evaluate an individual prediction skill, the mean square error of individual prediction (MSEIP) is used for all leads up to 12 months, as defined in Tang et al. (2008a, b),

$$MSEIP_i = \frac{1}{12} \sum_{t=1}^{t=12} \left(T_i^p(t) - T_i^o(t) \right)^2 \quad (3.3)$$

3.2.2 Cross-Wavelet Analyses

The Cross-wavelet transform (XWT) method is used for examining relationships between two time series in time-frequency space (e.g., Grinsted et al. 2004). From the XWT analysis, the common power and relative phase can be revealed. The phase differences between two variables are depicted by the direction of a vector, with in-phase pointing right, anti-phase pointing left, and the first variable leading the second by 90° pointing straight down. In this study, a continuous XWT technique with the Morlet wavelet as the mother function was applied. Monte Carlo methods are used to assess the statistical significance against a red noise background. The standard software package of cross-wavelet transform is available online*. Further details on XWT analysis can be found in Grinsted et al. (2004) and Torrence and Compo (1998).

* <http://www.pol.ac.uk/home/research/waveletcoherence>

3.3. The Singular Value and ENSO Predictability

As a potential predictability measure, the optimal perturbation growth rate ($S1$) presumably has an inverse relationship to the actual model prediction skill, namely, when $S1$ is large, the predictability is low and vice versa. Such a perception has been applied in studying potential predictability of ENSO using the theory of optimal perturbation growth (e. g., Moore and Kleeman 1998; Moore et al. 2006; Tang et al. 2006). However, the relationship between $S1$ and the actual prediction skill measures such as the anomaly correlation (R) and mean square error (MSE) to date has not been well examined due to a lack of long-term retrospective ENSO predictions and corresponding SV analysis, as discussed in the introduction. Different from previous SV studies, we will focus on discussing relationship for individual forecasts rather than an overall feature, which offers useful potential metric in estimating the performance of a forecast when verification data is absent. In the next section, the relationships between singular value and actual prediction skill metrics will be investigated at various time scales for the period of 148 years from 1856 to 2003.

3.3.1 $S1$ –Predictability Relationship on the Decadal/Interdecadal Time Scale

Firstly, we examine the relationship between the perturbation growth rate $S1$ and the ENSO prediction skills at the decadal/interdecadal time scales. All the skill measures presented in Section 3.2 are used, including anomaly correlation (R), MSE, and MSEIP. The correlation preliminarily evaluates the phase differences between the forecasts and observations, while the MSE and MSEIP quantify the amplitude departure between the forecasts and the observations. Unless otherwise indicated, the predicted and observed Nino3.4 SSTA indices are used to

evaluate these actual prediction skills in this study. The S1 was calculated with the optimal period of 9-months. As found in chapter 2, the fastest perturbation growth rate (maximum S1) occurs at a 9-month lead in the LEDO5 model. Correspondingly, the prediction correlation skill and MSE skill vary slowly with lead time after 9-month leads (Chen et al., 2004; Chen and Cane, 2008). This motivates us to choose the S1 of 9-month lead in the following discussions.

To examine the relationship of S1 to predictability on the interdecadal time scales, a running window of 25-yr was applied to the S1 and the actual prediction skill measures, namely that, they were evaluated at each window of 25 years, starting from January 1856 and moving forward by 1 month each time until December 2003. Since R and MSE are a function of lead time, their values averaged over lead times of 1-12 months are presented in Fig. 3.1. As can be seen in Fig. 3.1, on the interdecadal time scale over the 148 years, there is an inverse relationship between the S1 and the correlation skill (R) and an in-phase relationship between S1 and the MSE. These relationships are consistent with the conventional concept of S1 and predictability, namely, when the S1 is small, prediction skill was good, i.e., high correlation skill R and small MSE-based skill; whereas when the S1 was large, the opposite situation occurs. Note that the averaged MSEIP over a running window is equivalent to the averaged MSE over all lead times, as suggested by Eq. (3.2) and (3.3).

The running mean method used above may not be able to present objectively a full spectrum of the relationship between S1 and predictability; for example, the relationship is probably sensitive to the length of running window. To explore the decadal/interdecadal relationships, further we extract low-frequency components using the fast Fourier transform (FFT) filter. Shown in Fig. 3.2a are low-frequency components of S1 and the MSEIP, obtained

by a 10-yr low-pass FFT filter. Generally, Fig. 3.2a confirms the in-phase relationship in Fig. 3.1b, with a significant positive correlation coefficient of 0.4 over the 148 yrs between the S1 and MSEIP. A further scrutiny to Fig. 3.2a reveals that the in-phase relationship has decadal/interdecadal variation. Fig. 3.2b shows the correlation coefficient between the filtered S1 and MSEIP, computed using the running window of 25 yrs over the 148 yrs. As can be seen, the in-phase relationship between S1 and MSEIP was much stronger during the late 19th and 20th centuries than during the periods from 1910-1920 and 1940-1955. In following discussions, we will see that the in-phase relationship between S1 and predictability is most probably due to decadal variation in ENSO signals.

3.3.2 S1 – Predictability Relationship on Interannual Time Scales

In the proceeding section, an in-phase relationship was found between the S1 and MSE skill metric at long time scales greater than decade. A further analysis explores whether such an in-phase relationship exists at interannual time scales and for individual forecast cases. Shown in Fig. 3.3 is the scatter plot of S1 against MSEIP, where a 2-7 yr FFT filter has been applied to both variables to extract their interannual variability. Fig. 3.3 indicates large uncertainties in the relationship between S1 and MSEIP, suggesting that, on the interannual time scales, the optimal error growth rate S1 might not be a good indicator of actual model skill.

3.3.3 S1 – Predictability Relationship on All Time Scales

For all time scales ranging from seasons to decades, the relationship between S1 and predictability is measured using all original samples without filtering, as shown in Fig. 3.4 and the second column of Table 3.1. Here, MSE and R were evaluated in a period as a function of

lead time, making them unavailable in Table 3.1. As shown in Fig. 3.4, a large uncertainty exists in the relationship between S1 and MSEIP, with a low correlation value of 0.16.

In summary, the relationship between S1 and predictability is complex, dependent on time scales and the target of evaluation. At decadal time scales, S1 has an in-phase relationship to MSE and an inverse relationship to correlation skill; whereas at interannual time scales and for individual forecasts, the relationships between S1 and prediction skill measures have larger uncertainties. Thus S1 might not be the best indicator of predictability. In next section, we will further explore S1 and propose a better measure for quantifying potential predictability.

3.4. The Linear/Nonlinear Perturbation Growth Rates and the Actual Predictability

As analyzed in the proceeding section, there are significant uncertainties in the relationship between S1 and predictability at interannual time scales and for individual initial conditions. Conceptually, a good relationship between them should be expected since S1 quantitatively measure the fastest error growth. However, the potential predictability measure S1 is the fastest error growth rate, which might not always indicate the actual predictability in the actual forecasts. Thus, it is interesting to explore additional possible reasons responsible for the uncertainties of S1 and actual predictability, in particular, to identify better measures of potential predictability than S1.

Practically, the perturbation growth, denoted by δS , can be decomposed into the

perturbation growth due to the linear heating (LH), δL , and that due to the nonlinear heating⁺ (NH), δN , namely;

$$\delta S = \delta L + \delta N \quad (3.4)$$

δS , δL , and δN are the final perturbation growth at the lead time of 9-month, obtained by the TLM with the SV1 as the initial perturbations. In chapter 2, we found that there is a strong inverse relationship (with a correlation coefficient of -0.81) between δL and δN in the central and eastern Pacific; the linear perturbation growth δL is about twice as large as δN in amplitude, and δN is always negative whereas δL is positive in most time. Thus, the strong inverse relationship between linear and nonlinear perturbation growth can be approximately depicted as below

$$\delta N \approx b \delta L, \quad (3.5)$$

where b is a constant value with $-1 < b < 0$.

Applying the L-2 norm on (3.5) and (3.4), respectively, we have

$$\|\delta N\| \approx |b| \cdot \|\delta L\| = -b \|\delta L\|. \quad (3.6)$$

$$\|\delta S\| = \|\delta L + \delta N\| = \|\delta L + b \delta L\| + \Delta S = (1 + b) \|\delta L\| + \Delta S = \|\delta L\| + b \|\delta L\| + \Delta S \quad (3.7)$$

In (3.7), the ΔS is the residual term arising from the approximations in (3.5) and (3.6),

⁺ Note that the nonlinear heating term has been linearized in the TLM, thus the nonlinear perturbation growth used in this paper, unless otherwise indicated, means the perturbation growth due to the nonlinear heating term linearized in the TLM.

representing the perturbation growth due to the interaction of linear and nonlinear heating. The value of $1+b$ is always positive due to of the condition $-1 < b < 0$. Thus,

$$\|\delta S\| = \|\delta L\| - \|\delta N\| + \Delta S \quad (3.8)$$

Based on the definition of perturbation growth rate (singular value), the total optimal perturbation growth rate $S1$ can be decomposed into the contribution of LH (referred to as linear growth rate $L1$) and the contribution of NH (referred to as nonlinear growth rate $N1$), given by (3.8), namely,

$$S1 = L1 - N1 + \Delta S = L1 + NN1 + \Delta S, \quad (3.9)$$

where $NN1 = -N1$, a negative value representing a negative/offsetting contribution of NH to $S1$. In following analysis, we will find that the ΔS is small compared with $L1$ and $NN1$.

Shown in Fig. 3.5a-c are scatter plots of $L1$, $NN1$ and ΔS against $S1$ for the period from 1856-2003; where ΔS is obtained by $S1 - (L1 + NN1)$. As can be seen, the contribution of ΔS to $S1$ is rather small (ranging from 0 to 5 with the mean value of 2.0 in Fig. 3.5c), and $S1$ is mainly determined by the sum of the perturbation growth rates $L1$ and $NN1$. Fig. 3.5a-c indicates that $L1$ and $NN1$ have an offset effect or an opposite contribution on $S1$, i.e., a positive relationship between $L1$ and $S1$ in contrast to an inverse relationship between $NN1$ and $S1$. Such an offsetting effect might be a preliminary reason why $S1$ is not a good indicator of actual prediction skill as found above. In other words, either $L1$ or $NN1$ might be expected to have a better relationship with actual prediction skills than $S1$.

A strong anti-correlation between $L1$ and $NN1$ stems from the underlying dynamical

processes (i.e. linear heating LH and nonlinear heating NH) as argued above. To illustrate the strong inverse relationship between LH and NH, a scatter plot of NH against LH is given in Fig. 3.5d, where the LH/NH is the averaged linear/nonlinear heating at the NINO3.4 region over the lead times from 1 to 9 months for individual forecasts. A strong inverse correlation between LH and NH is visible with a significant correlation coefficient value of -0.89. As seen in Fig. 3.5d, the LH mainly has a warming effect in about 72% of forecasts for the 148 yrs, whereas the NH always contributes to a cooling effect. The cooling effect of the NH becomes stronger as the warming effect of LH increases, leading to a strong offsetting effect between LH and NH in most cases. Due to the strong offsetting effect in the underlying dynamical processes, the total heating (LH+NH) has a poor relationship with the total error growth S1 with a small correlation coefficient of 0.13.

Table 3.1 shows the correlation coefficients between the actual prediction skill MSEIP and potential predictability measure L1/NN1 (the third and fourth column). As can be seen, both L1 and NN1 have stronger relationships to prediction skills than S1, indicating NN1 or L1 to be a better measure of potential predictability in the ZC model. A comparison between these correlation coefficients reveals that the L1 and NN1 have opposite relationships to predictability, as displayed by a positive correlation coefficient for L1 but a negative value for NN1. This offset effect might explain well why there is a relatively small correlation value between S1 and predictability as shown in Table 3.1. From the sign of correlation coefficients, one can infer that the positive S1 – MSEIP relationship is mainly determined by the positive L1 – MSEIP relationship.

Shown in Fig. 3.6 are scatter plots of L1 and NN1 against the prediction skill MSEIP. Like

Fig. 3.3, a 2-7 yr FFT filter was applied here for each variable to address interannual time scales. In contrast to the large uncertainties in the relationship of S1 – predictability as shown in Fig. 3.3, Fig. 3.6 shows visible relationship of L1/NN1 – predictability. The correlation coefficients between L1/NN1 and prediction skills are all statistically significant at the confidence level of 95%, as shown in the upper-left corner of each panel. An opposite relationship between L1-MSEIP and NN1-MSEIP can be observed in Fig. 3.6, showing an offset effect of L1 and NN1 on predictability, as found earlier in Table 3.1. Thus, Fig. 3.6 explains the large uncertainties in S1 – predictability in Fig. 3.3, and also indicates that either L1 or NN1 is a better measure of potential predictability than S1.

A further analysis is placed on the relationship between the LH/NH and actual prediction skill. Table 3.1 includes the correlation coefficients between the averaged LH/NH over the optimal period of 9 months and the actual prediction skill. As shown in the fifth and sixth columns of Table 3.1, LH and NH significantly correlate with MSEIP, namely, when LH/NH is strong, the MSEIP skill is large and vice versa. Again, the NH has always a cooling effect as aforementioned, thus the negative sign of its correlation coefficient in Table 3.1 imply the link of stronger cooling and the larger MSEIP. This explains the relationships between the L1/NN1 and the prediction skill. These results are in agreement with our previous findings that both the model linear/nonlinear perturbation growth rate and the linear/nonlinear heating term are controlled by the underlying linear/nonlinear processes, respectively.

To illustrate the time-scale-dependent characteristics of the relationship between perturbation growth rates and MSEIP, we performed cross-wavelet analyses for them as shown in Fig. 3.7. The in-phase relationship is presented by arrows pointing right, whereas the anti-

phase (or inverse) relationship is displayed by arrows pointing left. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. Several features can be revealed by Fig. 3.7. First, the scale-dependent feature of the S1 – MSEIP relationships is seen in Fig. 3.7a. At longer time scales >10 yrs, a strong in-phase relationship is displayed in the S1 – MSEIP for the periods of 1860-1940 and 1970-2000 (Fig. 3.7a), which is in agreement with the decadal variations of correlation coefficients in Fig. 3.2b. At shorter time scales <10 yrs, wavelet analysis reveals additional scale-dependent relationships. For example, at 6-10 yrs time scale, anti-phase relationships are shown in two time periods of 1890-1910 and 1970-1980, which are opposite to the in-phase relationships displayed at decadal/interdecadal time scales. For 2-6 yrs time scale, in-phase relationships occurred again but confined in 1860-1900 and around 1960. This scale-dependent relationship is consistent with the results shown in section 3.3. Second, at all time scales, the S1 – MSEIP relationship (Fig. 3.7a) looks more like the L1 – MSEIP relationship (Fig. 3.7b) than the NN1 – MSEIP (Fig. 3.7c). This similarity is because that the contribution of LH to SSTA is about twice as much as the NH (In chapter 2), thereby the original S1 – MSEIP relationship is mainly determined by the L1 – MSEIP relationship. Third, at interannual time scales, the L1 (NN1) shows a more frequently consistent in-phase (anti-phase) relationship with MSEIP, suggesting L1 or NN1 is a better measure than S1. Furthermore, the NN1 – MSEIP relationship (Fig. 3.7c) is consistently inverse for almost over all the time scales. It does not have the scale-dependent feature like that in S1 – skill (Fig. 3.7b) and L1 – skill (Fig. 3.7b), where in-phase and anti-phase relationship change alternatively from time to time. This unique feature of the NN1 – skill relationship suggests that NN1 is a more reliable measure of potential predictability. It should be noted that the NH has much smaller contribution to SSTA, but NN1 has a consistently significant anti-phase relationship

with MSEIP skill at all time scales, suggesting that a strong negative perturbation growth is related to a large MSEIP. Another feature shown in Fig. 3.7 is that, at interannual time scale, the L1 and NN1 brings a strong offsetting effect on MSEIP (opposite arrow direction) during the period from 1910-1960, leading to a large uncertainty in the relationship between S1 and MSEIP as shown in Fig. 3.3.

3.5. ENSO Signals, the Optimal Error Growth Rates, and Predictability

It has been suggested in many recent studies that ENSO predictability is strongly associated with signal components present in initial fields (e.g., Peng and Kumar 2005; Tang et al. 2005, 2008a; Moore et al. 2006). Often, a stronger ENSO event is easier to predict than a neutral event. At the decadal/interdecadal time scales, Tang et al. (2008a) compared retrospective ENSO predictions of 120 years from three models and found that, at the decadal/interdecadal time scales, high correlation skills often occurred at the time periods with strong ENSO events whereas low correlation skills occurred at weak ENSO periods. The positive relationship between ENSO signals and the correlation skill was explained in Kleeman (2002), Tang et al. (2005), and Tang (2008a), using information theory. They argued that the extra information provided by the forecast, called prediction utility, is highly associated with the signals present in the initial conditions. As the ENSO signal is stronger, more extra information will be produced compared with the climatological forecast, which leads to a more skillful and reliable forecast. However, the ENSO signal and predictability at shorter time scales, i.e., interannual time scales, has not been well addressed. In this section, we will examine relationships between the ENSO signal and the optimal growth rates at time scales ranging from interannual time scale to decadal/interdecadal time scales.

An inverse relationship has been suggested between the optimal perturbation growth and the intensity of ENSO variability, in some recent SV and breeding vector analyses: a small perturbation growth rate $S1$ often occurs during an ENSO peak phase, and the larger perturbation growth rate $S1$ appears in the neutral and onset/breakdown stages of ENSO (Chen et al. 1997; Xue et al. 1997a; Tang et al. 2005; Cai et al. 2003; Zhou et al. 2008). These works identified the inverse relationship either using a comparison of the maximum $S1$ against the intensity of ENSO variability or using an analytical solution of the delay oscillator model.

A metric to measure the intensity of the ENSO signal should be defined. Tang et al. (2008a) proposed three measures to quantify the intensity of ENSO over a time period including: i) the variance of NINO3.4 SSTA index, ii) the variance of the first EOF mode, and iii) total spectrum power at frequencies of 2-5 years. Tang et al. (2008a) shows that the three measures produce similar decadal/interdecadal variation of ENSO signal. In Part I of this work (In chapter 2); ENSO signal was defined by the absolute value of NINO3.4 SSTA index. In the present study, we use the same definition to measure the intensity of ENSO signal to be consistent with Part I.

The relationships between $S1/NN1/L1$ and ENSO signals are displayed in Fig. 3.8 using cross-wavelet analyses. At the decadal/interdecadal time scales, the $L1$ and $NN1$ have stronger relationships to ENSO signals than the $S1$, especially for the period from 1880 to 1910 and around 1980. This is especially true for the $NN1$ – signal relationship which holds for almost all the periods from 1880-1980. At the decadal/interdecadal time scales, the $S1$ – signal relationship in Fig. 3.8a is determined by both $L1$ and $NN1$. As can be seen, the anti-phase $NN1$ – signal relation cancels the in-phase $L1$ – signal relationship completely in 1860-1900 and partly in the 1980s and later. Therefore, both $L1$ and $NN1$ have important contributions to the

original S1 – signal relationship at decadal/interdecadal time scales. These features are furthered revealed in plots of variations in S1/NN1/L1 against signal as shown in Fig. 3.9.

On the interannual time scales, the relationship between S1 and ENSO signals is not clear in Fig. 3.8a. The in-phase and anti-phase relationships occur randomly from decade to decade. On the other hand, for most periods during the 148 years, the L1–signal and NN1–signal show consistently good in-phase and anti-phase relationship, namely, a strong ENSO signal is associated with a large L1 (small NN1) while a weak ENSO signal corresponds with a small L1 (large NN1). Especially, the NN1 – signal relationship is most significant during the 148 years at both decadal and interannual time scales.

The relatively good relationship between NN1/L1 and ENSO signals can be further demonstrated in the plots of variations in NN1/L1/S1 against ENSO signals. Shown in Fig. 3.9 are these variations at decadal time scale. As can be seen, a much better relationship between NN1 – ENSO signals can be identified, which explains the importance of nonlinear heating in ENSO variability and predictability as found in other studies (e.g., Tang and Deng 2009). For interannual time scale, we also found that the ENSO signal is more related to NN1 than to others (not shown).

3.6. Conclusion

In this study, we investigated ENSO predictability using the optimal perturbation growth and long-term retrospective hindcasts using the ZC model. Emphasis was placed on exploring the relationship between potential predictability measured by the optimal perturbation growth rates and actual hindcast skill for long-period from 1856 – 2003. A good measure of potential

predictability is useful practically because it can estimate the prediction skill without using the observations, and offer a practical means of estimating the confidence level of an individual prediction.

To find the best measure of potential predictability, three metrics obtained from SV analyses in chapter 2 have been examined at different time scales, including the leading singular value S1, the linear (L1) and linearized nonlinear (N1) components of S1. The L1 and N1 reflect the optimal perturbation growth of the linear and nonlinear heating terms in the SST governing equation of the ZC model. The measures of actual prediction skill include correlation coefficient, MSE, and mean square error of individual prediction (MSEIP). Generally, at decadal/interdecadal time scales, our findings from the long-period analysis of 148 years confirmed the theoretical perception that S1 has an inverse relationship with correlation-based skill, and a positive relationship with MSE-based skills. However, at shorter time scales, e.g., interannual time scales, and for individual forecast cases, there are large uncertainties in the relationship between S1 and actual prediction skills, which prevents the S1 from being a good measure of potential predictability.

Several reasons are probably responsible for the small correlation between S1 and actual predictability. First, S1 is a collective error growth jointly contributed by the linear and nonlinear processes. A strong inverse relationship between LH and NH might cause an unrealistic offsetting contribution to S1, as indicated by strong anti-correlation between L1 and NN1, biasing the relationship between S1 and prediction skill. Instead, L1 or NN1, removing the offsetting influence, might better characterize the relationship between potential skill and actual skill. Second, S1/L1/NN1 is a potential measure, and represents the optimal/fastest error

growth rate but such an extreme situation does not always happen in the realistic forecasts. Therefore, even under the perfect model scenario, they still may not have a very good relationship with the actual predictability. S1 represents a more extreme situation than NN1 and L1 since it contains L1, NN1, probably leading to worse relationship to actual skill. Third, the relationship between potential and actual skill is also influenced by model bias inherent to model internal dynamics and physical processes. The model is always not perfect. The S1 involves more physical processes than either L1 or NN1, thus the model bias can more easily impact S1 than L1/NN1, more biasing the relationship between potential predictability and actual prediction skill.

An important finding in this work is that the linear/nonlinear perturbation growth rate L1 and NN1 are better measures of potential predictability than the optimal perturbation growth rate S1 in terms of the capability of estimating the actual prediction skills. Among the three potential measures, NN1 has a consistent relationship with actual prediction skills for all time scales. Uncertainty in the relationship between S1 and prediction skill measure is due to an offsetting effect of linear heating and nonlinear heating on the optimal perturbation growth, causing an opposite relationship between L1-predictability and NN1-predictability.

A practical application of this study is to use L1 and NN1 to characterize potential predictability. It was also found that the residual term in Eq. (3.9) has small contribution to the S1, allowing to use the sum of L1 and NN1 to replace S1. The analysis of L1 and NN1 can be applied to all time scales and is suitable for individual cases and overall features. It should be noted that a high correlation skill and a large MSE value can occur simultaneously, namely one prediction is good in phase but poor in magnitude. This is most probably due to the nature of

prediction target whose variance is large. It is well recognized that strong El Niño events might be easier to predict than normal events but the prediction errors in amplitude often are larger for strong ENSO events. Thus, it might be necessary to draw conclusions and summarize findings from the two different predictability measures.

The perturbation growth rate $L1/NN1$ depends on the nature of initial conditions and the internal dynamical processes (i.e., linear and nonlinear heating). The latter often controls the intensity of ENSO variability. Due to the offsetting effect of linear and nonlinear heating on ENSO variability and the time-scale dependent nature of these dynamical processes, the relationship between $S1$ and ENSO signals depends on both the time periods and time scale (e.g., Fig. 3.8). For example, an inverse relationship can be identified on the interannual time scales over the recent decades (after 1960s), consistent with those documented in previous BV and SV studies. However this inverse relationship does not hold well for other periods and for other time scales. In contrast to the uncertain $S1$ -signal relationship, the $NN1$ shows a consistent inverse relationship with ENSO signals for all periods and time scales.

Several cautions should be borne in mind. First, the SV analyses and retrospective hindcasts are often model-dependent, suggesting that the results and conclusion drawn from this work might not be generalized. More models are required to fully generalize these conclusions. Second, some physical processes are either simplified or missing in the ZC model. For example, stochastic atmospheric noise is not considered in this model. Stochastic forcing has been thought to be a main source to limit ENSO predictability. Thus, the predictive skill shown in the ZC model might be a lower bound of ENSO actual prediction skill (Chen and Cane 2008). Third, the total nonlinear heating NH always contributes to a cooling effect in the ZC model,

which is opposite to the observation as discussed in An and Jin (2004) where the vertical nonlinear warming dominates the total nonlinear heating term. This is due mainly to model unrealistic simulation of the zonal current anomaly. Thus, some results found in this work may be model dependent. However, the unrealistic simulation of NH is common in current ENSO prediction models. Comparing nonlinear heating terms in ten coupled models reveals that only one model gave the correct simulation of NH and others fail to represent both the location and strength or even the sign of the nonlinear vertical warming (An et al. 2005). Fourth, the results and conclusions in this study might be also dependent on the metrics of actual prediction skill. In this study, we explored ENSO predictability using correlation-based and MSE-based measures, especially MSEIP. When the chosen metrics have been widely used in the field of predictability study, they might not be able completely characterize all properties of predictability. Finally, we used a running window of 25-yr to analyze interdecadal variations in predictability and other variables. The window length of 25-yr was arbitrary and subjective although several sensitivity experiments with different window lengths did not essentially change the aforementioned results. These concerns need to be addressed through more comprehensive analyses.

Nevertheless, this work explored ENSO statistical predictability over the past 148 years, providing insights on ENSO predictability, especially offering a practical means to estimate the confidence level for individual forecasts for the ZC model. An investigation of individual error growth rates i.e. the linear perturbation growth $L1$ and the nonlinear perturbation growth $N1$ from their controlling processes (the underlying linear and nonlinear advections) offer the better potential measures for ENSO predictability. Since the perturbation growth $L1$ and $N1$ are determined by the underlying linear and nonlinear dynamical processes respectively, these

processes are fundamental reasons that contribute to the strong relationships of signal/perturbation growth and ENSO predictability. For example, the relationship between N1 and forecast skill probably is the result of two known relationships: i) relation between ENSO magnitude and forecast skill, ii) relation between the nonlinear heating and ENSO magnitude.

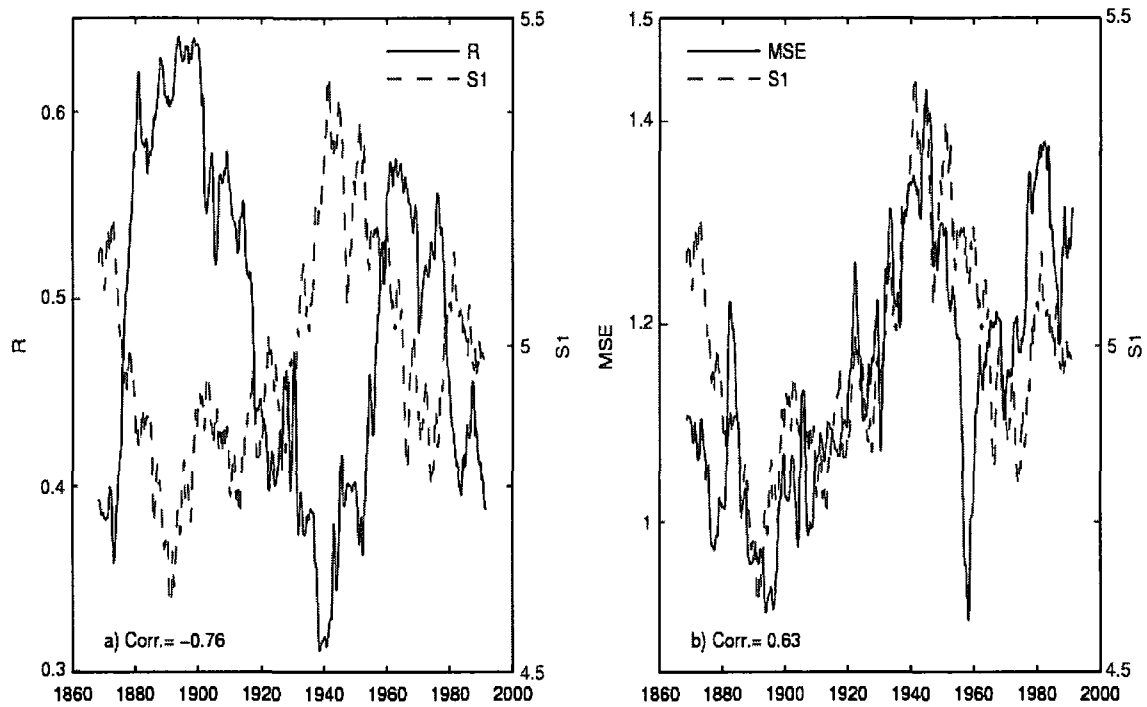


Fig. 3.1 Interdecadal variations of a) anomaly correlation coefficient (R) and the singular value (S1); b) MSE and S1. A 25-yr running window was applied on all data at each month from Jan. 1856 to Dec. 2003. MSE measures are averaged over lead times of 1-12 months.

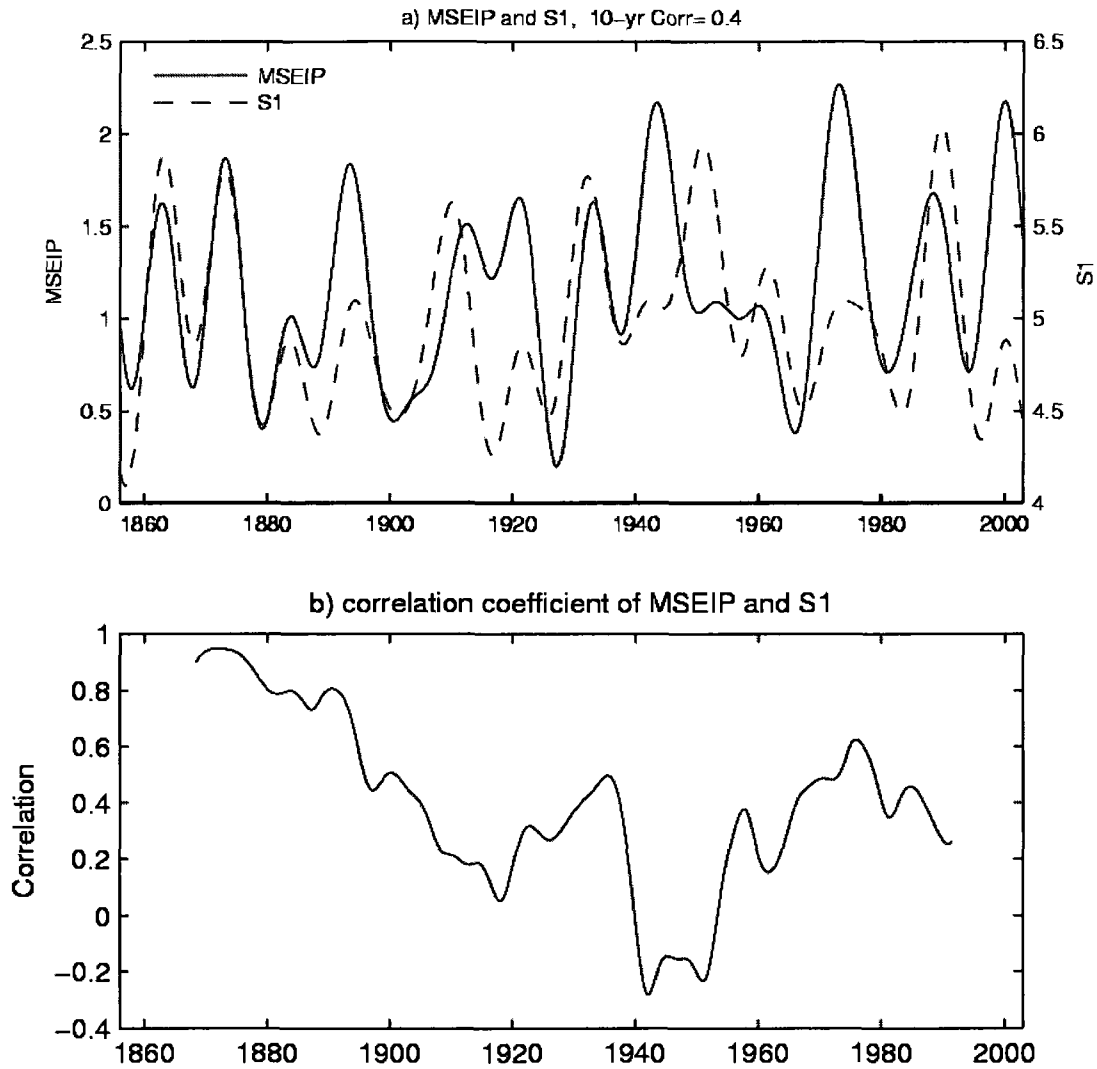


Fig. 3.2 a) Decadal/interdecadal variations of MSEIP and singular value (S1). A 10-yr low-pass FFT filter method was applied on these skill measures. b) Temporal variations of the correlation coefficient between S1 and MSEIP over the 148 years, correlation coefficients were calculated in a 25-year running window.

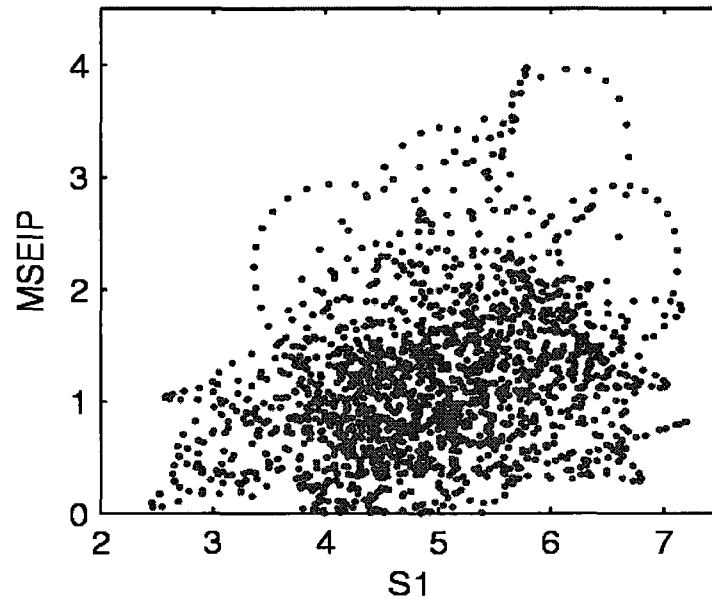


Fig. 3.3 The relationships between the singular value S1 and the actual predictability measures at interannual time scales using a 2-7-yr FFT filter. S1 against MSEIP.

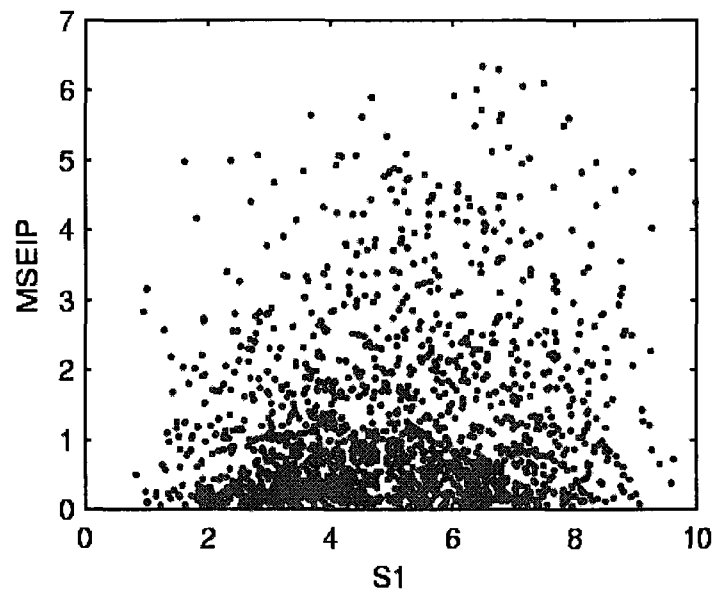


Fig. 3.4 Same as Fig. 3.3 but for all time scales without using an FFT filter.

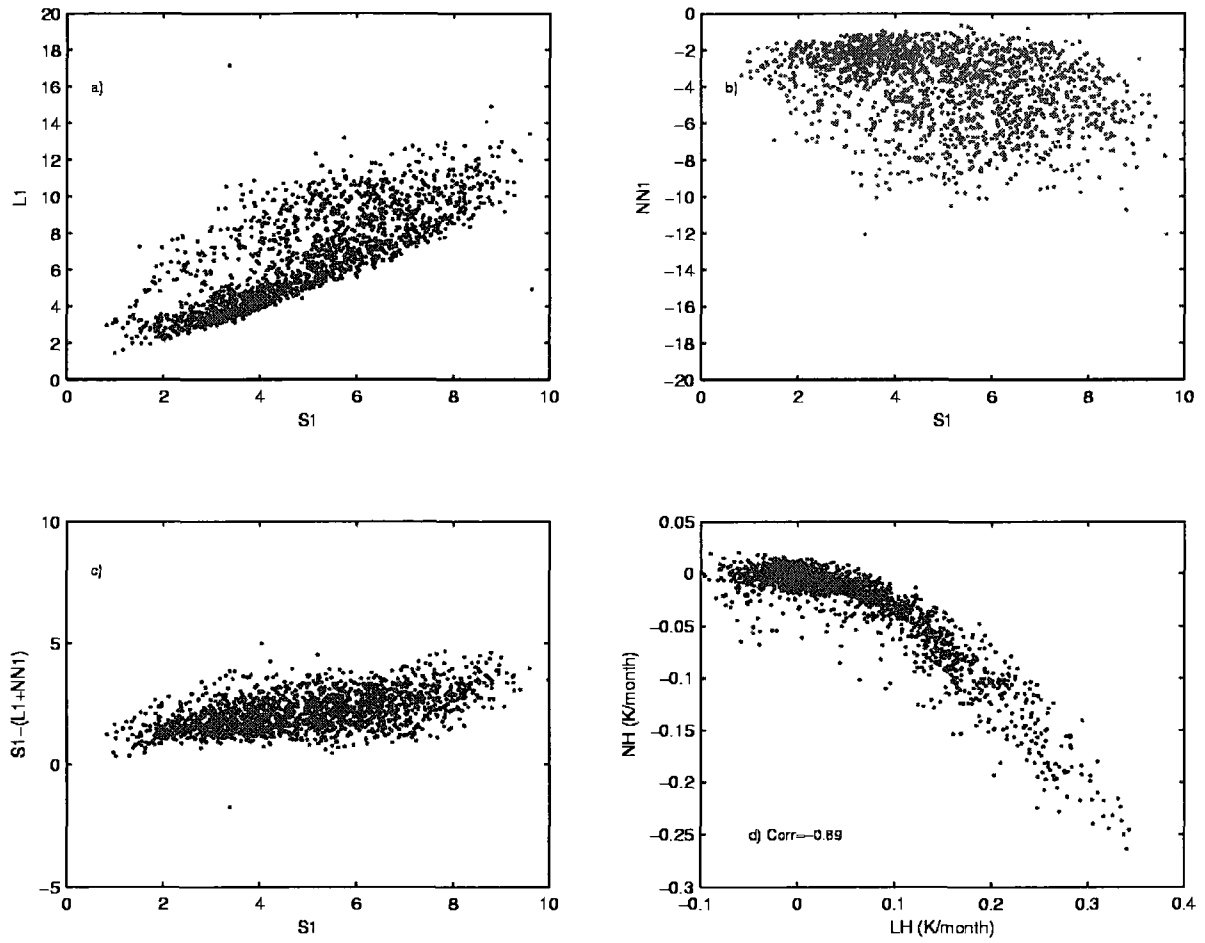


Fig. 3.5 Scatter plots give the relationships of the linear ($L1$), nonlinear ($NN1$), and the total ($S1$) perturbation growth rates. a) $L1$ against $S1$. b) $NN1$ against $S1$. c) $S1 - (L1 + NN1)$ against $S1$. d) Linear heating LH and nonlinear heating NH (K/month)

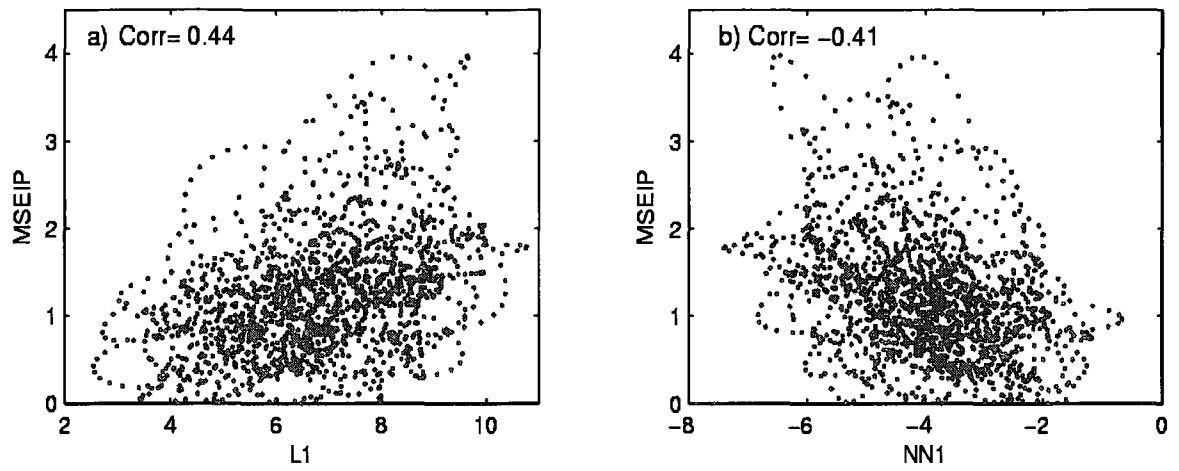


Fig. 3.6 The relationships between the linear/nonlinear singular values and the prediction skill MSEIP at interannual time scales using a 2-7-yr FFT filter. a) L1 against MSEIP; b) NN1 against MSEIP.

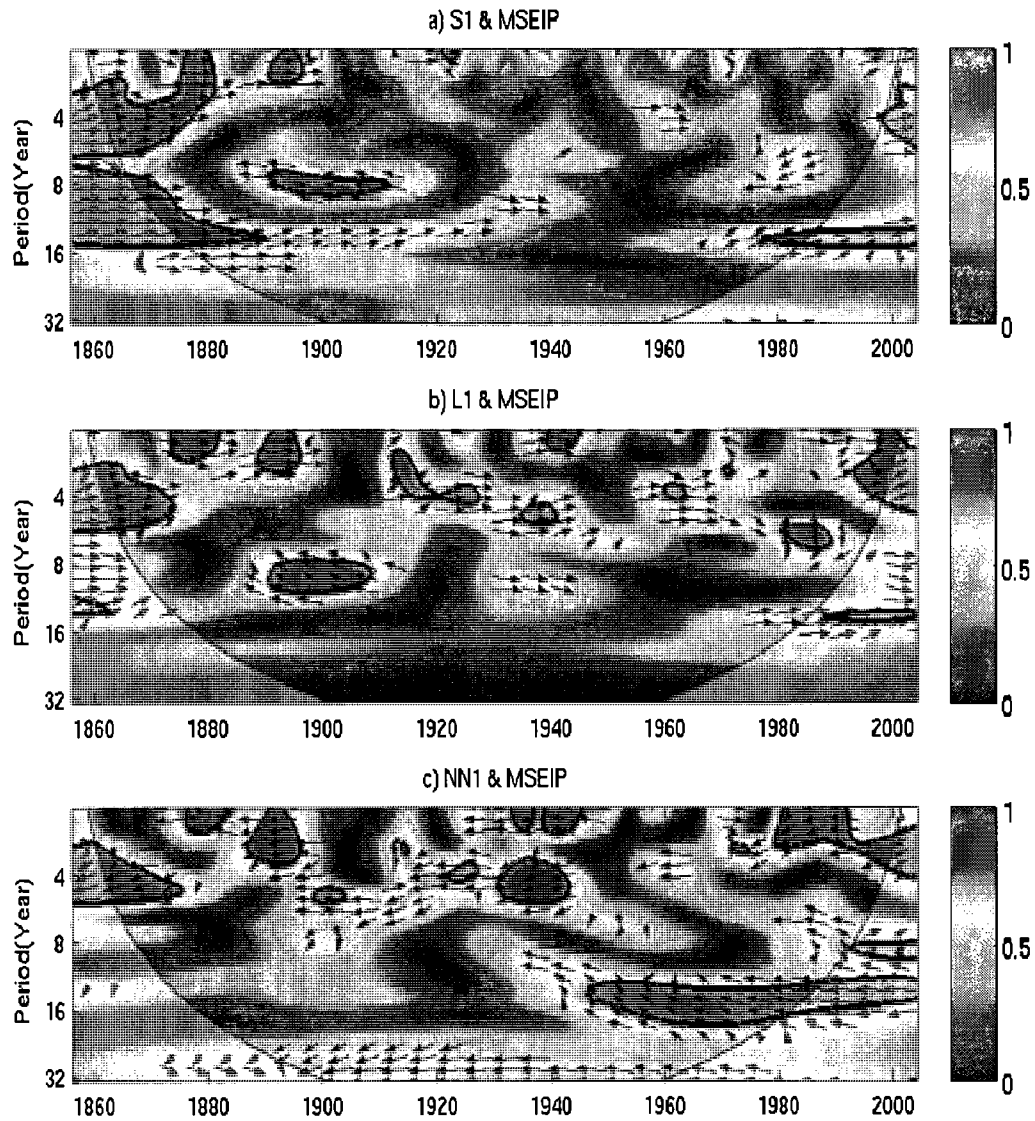


Fig. 3.7. The cross-wavelet analysis for the singular values S1/L1/NN1 and actual prediction skill MSEIP. a) S1 and MSEIP; b) L1 and MSEIP; c) NN1 and MSEIP. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left, and singular values leading skills by 90° pointing straight down. (A 2-yr FFT filter was applied on all data before performing the cross-wavelet analyses).

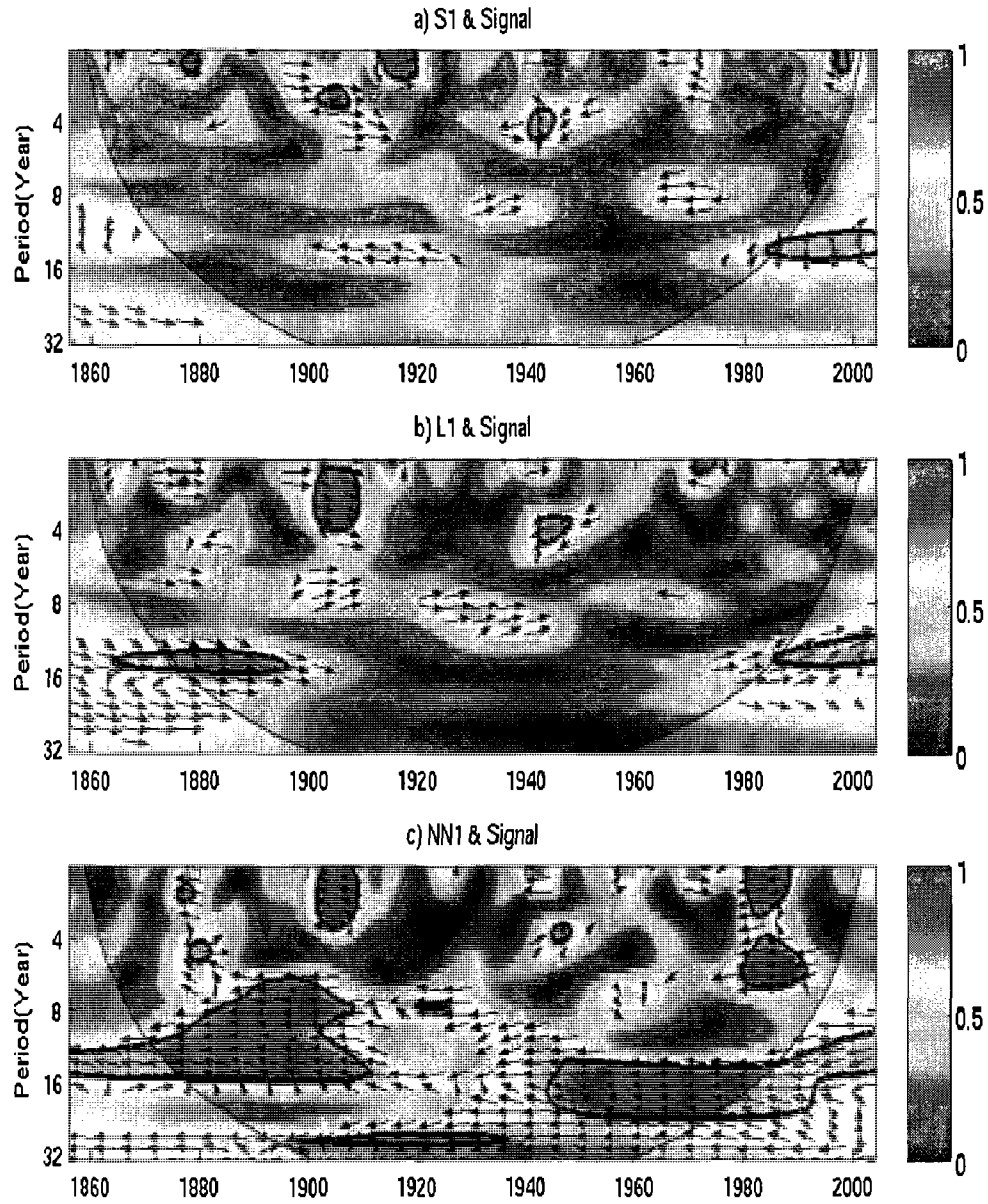


Fig. 3.8 The cross-wavelet analysis for ENSO signal (NINO3.4) and the singular values S1, L1, and NN1.

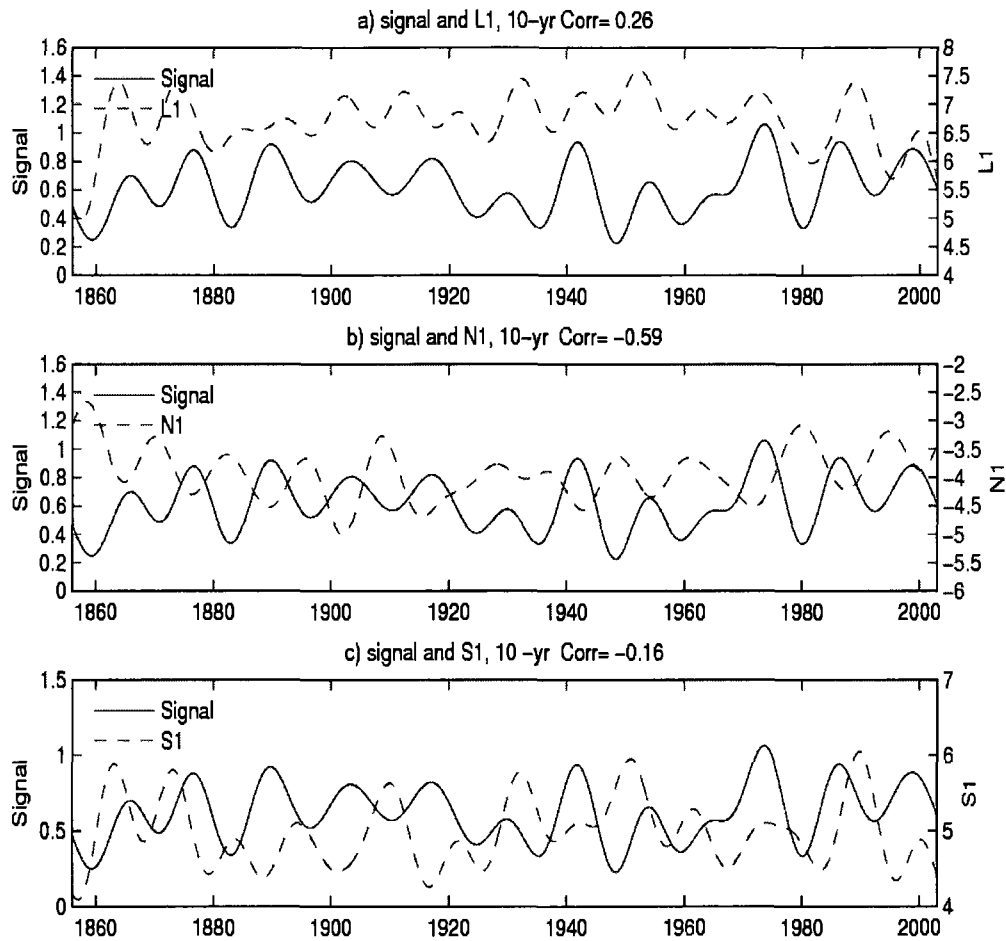


Fig. 3.9 Decadal/interdecadal variations of ENSO signal ($|NINO3.4|$; the solid line) and perturbation growth rates (dash lines): a) the linear perturbation grow rate (L1); b) the linearized nonlinear perturbation growth rate (NN1); c) the total perturbation growth rate S1. A 10-yr low-pass filter has been applied.

Table 3.1 Correlation coefficients of potential predictability measures and actual predictability measure. The actual predictability measure is MSEIP, whereas the potential predictability measures include the leading singular value (S1), the linear perturbation growth rate (L1) and the nonlinear perturbation growth rate (NN1). The LH and NH represent the linear heating and nonlinear heating items in SST governing equation, averaged over the NINO3.4 region and over the optimal period of 12 months.

	S1	L1	NN1	LH	NH
MSEIP	0.16	0.33	-0.36	0.58	-0.51

Chapter 4: Ensemble Construction and Verification of the Probabilistic ENSO Prediction in the LDEO5 Model

Cheng Y, Tang Y, Jackson P, Chen D, and Deng Z (2009) Ensemble Construction and Verification of the Probabilistic ENSO Prediction in the LDEO5 Model (submitted to J. Climate on September 25, 2009, revised in January 2010).

4.1. Introduction

The loss of ENSO predictability in a numerical model generally depends on uncertainties due to i) errors in the initial conditions, ii) model errors, and iii) unexpected external stochastic noise (e.g., Moore and Kleeman 1998). These uncertainties develop during the forecast period as lead time increases, eventually rendering the forecast no better than climatology. As a response to the limitations imposed by these uncertainties, a more useful forecast strategy is to perform ensemble predictions and evaluate the uncertainties of the forecast system using probabilistic methods (Chen and Cane, 2008).

To perform an ensemble-based ENSO probabilistic forecast, the crucial issue is to design a reliable and high resolution ensemble prediction strategy that should include the major uncertainties of a forecast system. Many strategies have been used in the ensemble construction of weather forecasts and seasonal climate predictions. For example, some strategies are dynamically constrained methods such as breeding-vector (BV; Toth and Kalnay 1993), the ensemble transform (ET; an improved version of the BV; Bishop and Toth 1999; Wei et al. 2008), and the singular vector (SV; e.g., Lorenz 1965; Palmer 1993), they are used to optimally perturb the initial conditions for constructing ensemble forecasts. Other methods are also used to obtain the “best” initial conditions in ensemble constructions: the ensemble Kalman filter (ENKF; Evensen 1994; Evensen 2003), the ensemble transform Kalman filter (ETKF; Bishop 2001; Wang and Bishop 2003), and the perturbed observation (PO; Houtekamer and Derome 1995). Using the three-parameter Lorenz (1963) model, Anderson (1997) found that random perturbations produce more skillful ensembles than BV and SV. Houtekamer and Derome (1995) found little difference in the quality of the ensemble mean

forecasts between the BV, SV, and PO methods using a quasigeostrophic model. Hamill et al. (2000) compared BV, SV, and PO in a quasigeostrophic channel model. They found that the PO method is better than the BV and SV method. Descamps and Talagrand (2007) compared four strategies in the Lorenz (1963) model and a three-level atmospheric model and concluded that the relative performance, from best to worst, of these strategies was in the order $\text{EnKF} > \text{ETKF} > \text{BV} > \text{SV}$.

Significant progress has been made in using these optimal perturbations to study ENSO predictability as cited above. However these previous studies mainly focused on the optimal error growth of ENSO deterministic predictions. The impact of perturbation construction on the ensemble probabilistic predictions has not been well addressed, especially using long-term retrospective ensemble predictions over periods as long as 100 years. In this study, we will explore this issue using SV-based perturbation methods. So far, the SV method itself has not been well examined in the framework of ENSO ensemble probabilistic prediction. One reason is that the SV analysis needs tangent linear model (TLM), which is often technically difficult. Another reason is the lack of a long-term forcing data for initializing predictions, so that previous retrospective predictions were limited to a short period of 20-40 years, with a rather limited number of ENSO cycles. This may preclude statistically robust conclusions. Chen et al. (2004) used KAPLAN SSTA reanalysis data and the ZC model (LDEO5 version) to perform a 148 year hindcast experiment for the period of 1856-2003. They successfully predicted all of the prominent El Niño events during this period at lead times of up to two years, with the SST being the only data used for model initialization. Tang et al (2008a) further analyzed the interdecadal variation in ENSO prediction skill from 1881-2000 using multiple models. These retrospective ENSO predictions not only allow us to achieve a robust and stable study of

statistical predictability of ENSO but also demonstrated that the long-term SSTA data are of good quality. Recently, a fully physically-based TLM was constructed for the LDEO5 model, and singular vector analyses were performed for the 148 year period from 1856-2003 in Cheng et al. (2009). The long-term SVs obtained in Cheng et al. (2009) makes it possible to construct ensemble predictions with the LDEO5 model, so that the shape of the forecast probability density function (PDF) that describes the prediction uncertainty can be estimated, and the probabilistic nature of ENSO predictability can be explored.

Another issue is the role of stochastic atmospheric noise in ensemble ENSO predictions. It has been well recognized that stochastic atmospheric forcing associated with synoptic-to-intra-seasonal variability is critical in forming, developing and maintaining ENSO cycles (e.g., Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Eckert and Latif 1997; Blanke et al. 1997; Kirtman and Schopf 1998; Moore and Kleeman 1999; Thompson and Battisti 2000; Fluegel et al. 2004; Moore et al. 2006; Philip and van Oldenborgh 2009; Eisenman et al. 2005; Gebbie et al. 2007; Tziperman and Yu 2007; Zavala-Garay et al. 2003; Perez et al. 2005; Zhang et al. 2008). These studies consider that the high-frequency synoptic-scale atmospheric motion (i.e. weather events) and other high-frequency variations such as westerly wind bursts and the Madden-Julian oscillation (MJO) provide stochastic forcing to the ENSO modes and hence acting as a limit to the predictability. However, it is not very clear so far how the stochastic atmospheric noise impacts ENSO probabilistic predictions.

An important task associated with ensemble construction is to evaluate an ensemble-based probabilistic prediction system by probabilistic verification methods, from which the performance of the prediction system and the ensemble construction method can be quantitatively evaluated. Probabilistic verification is known as an important complement to

deterministic verification, which provides a useful and quantitative way to measure uncertainty (Palmer 2000; Kirtman 2003). In contrast with the traditional prediction skill measures such as anomaly correlation (R) skill and root mean squared error (RMSE) skill, the verification of an ensemble-based probabilistic forecast system focuses on measuring two properties: reliability and resolution, which are the two most important characteristics of a probabilistic forecast system (Toth et al. 2003). An introduction of these properties and the probabilistic verification methods will be described in section 4.4.

This study will introduce both initial condition uncertainty and additive stochastic atmospheric noise into the LDEO5 model and examine their impacts on ENSO probabilistic prediction. It is unrealistic to evaluate all ensemble construction methods available for ENSO probabilistic prediction, thus we focus on evaluating four methods, chosen based on previous studies as referred to above: (i) initial condition perturbation using the singular vector (SV) of SSTA (SV1_sst); (ii) realistic stochastic winds as a continuous external forcing during the forecast period (UV_realstoc); (iii) stochastic optimal winds (SO1_wind) as a continuous external forcing during the forecast period; (iv) a combination of the first method SV1_sst and the third method SO1_wind (SO1_wind+SV1_sst). Several probabilistic verification methods are used to evaluate the reliability and resolution of ensemble-based probabilistic ENSO predictions, including the reliability diagram (RD) and the Brier skill score (BSS), the ranked probability score (RPS), and the ranked probability skill score (RPSS). Emphasis is placed on assessing which ensemble construction method provides more reliable and higher resolution probabilistic ENSO predictions.

This Chapter is structured as follows: Section 4.2 briefly introduces the LDEO5 model, and the metrics of ensemble prediction skill. Section 4.3 discusses the four ensemble construction

methods used in this study. Section 4.4 gives the introduction of probabilistic forecast verification methods. Section 4.5 presents the ensemble prediction results followed by a conclusion and discussion in section 4.6.

4.2. Metrics for Ensemble Prediction Deterministic Skill

Several ensemble construction schemes are designed in this study, focusing on different aspects of uncertainties related to the predictability, i.e., the initial conditions and stochastically external forcing. These ensemble retrospective ENSO predictions were performed by perturbing SST or wind, or both, using a given method as described in section 4.3. The model is initialized by only the assimilation of SST every month for 1856-2003 from Chen et al. (2004), thus a total of 148 years \times 12 months/year (=1776) forecast initial conditions were obtained. From each initial time, an ensemble forecast was performed with the ensemble size (M) of 100, and for a period of 24 months. Thus, there are a total of 1776 months \times 100 members \times 24 months lead-time (=4262400) forecasts for the ensemble experiment of a given ensemble construction method.

In this study, we use the error of the ensemble mean ($RMSE_{EM}$) and ensemble spread to assess ensemble deterministic prediction skill, defined by,

$$SPREAD(i, t) = \sqrt{\frac{1}{M-1} \sum_{m=1}^M [T_i^p(m, t) - EM(i, t)]^2} \quad (4.1)$$

$$EM(i, t) = \frac{1}{M} \sum_{m=1}^{M=100} T_i^p(m, t) \quad (4.2)$$

where EM is the ensemble mean, a function of initial time i and lead time t . M is ensemble size,

i.e. 100 here. T is the index of Niño3.4 SSTA, the superscripts p and o denote predictions (forecasts) and observations respectively; N is the number of initial conditions used ($N=1776$).

$$SPRD(t) = \frac{1}{N} \sum_{i=1}^{i=N} SPREAD(i, t) \quad (4.3)$$

where the SPRD in (4.3) is the averaged ensemble spread over all the initial times, it is a function of lead-time (t) only.

4.3. Strategies of Ensemble Construction

4.3.1 Perturbation of Initial Condition with Singular vector (SV) of SSTA

In chapter 2, singular vector (SV) analysis was performed for the period 1856-2003 using the LDEO5 model. The leading singular vectors (SV1s), representing the optimal growth pattern of initial perturbations/errors, were obtained by perturbing the constructed tangent linear model (TLM) of the LDEO5 model. It was found that the first singular vectors of SSTA are dominated by a west-east dipole spanning most of the equatorial Pacific, with one center located in the east and the other in the central Pacific. The SV1s are less sensitive to initial conditions, i.e., are independent of seasons and decades. Thus, we will use the 148-year averaged SV1 of SSTA (denoted by SV1_sst) to perturb all initial conditions. As found in chapter 2, the fastest perturbation growth rate occurs at a 9-12-month lead in the LDEO5 model. Correspondingly, the prediction RMSE skill varies slowly with lead time after 12-month leads (Chen et al. 2004; Chen and Cane 2008). This motivates us to choose the SV1_sst of the 12-month lead in the following discussions. Note that the ensemble construction by two or more SV patterns does not show higher resolution or reliability than that constructed from

the SVI alone (not shown), thus only the SVI-based ensemble is used, so that we perturbed the initial model SST by the SVI_sst pattern. The construction of initial perturbation (Y) can be expressed by (4.4), where random numbers (X) were normalized, and α is a constant value controlling the perturbation magnitude, set to 0.25 here according to Karspeck et al. (2006).

$$Y = SVI_sst \times X \times \alpha \quad (4.4)$$

4.3.2 Realistic Stochastic Winds

In this study, we use two methods to generate the stochastic wind perturbations: high frequency (< 90 days) realistic winds, and stochastic optimal winds. The first of these, denoted by UV_realstoc, is our second ensemble construction strategy. A dataset of the atmospheric high frequency components were first obtained by applying a 3-month high-pass filter to the NCEP daily wind dataset from 1948-2000 (Deng and Tang 2008). This dataset, referred to as the noise dataset, realistically represents all possible temporal and spatial characteristics of atmospheric noise. Then, the atmospheric model (winds) is perturbed using the high frequency winds, randomly drawn from the noise dataset, at each model time step (10 days).

4.3.3 Stochastic Optimal perturbation

The spatial structure of initial perturbations has an important effect on the ensemble forecasts. The third method used for ensemble construction in this study is the stochastic optimal (SO) mode perturbation (Farrell and Ioannou 1993; Kleeman and Moore 1997; Moore et al. 2006; Tang et al. 2005; Tang et al. 2008b). Instead the realistic high frequency winds that might not generate optimal perturbation growth, we used the leading SO mode of winds (SO1_wind) to perturb the model through the entire forecast period. As discussed in Tang et al.

(2005; 2008b), for white noise in time, the SOs are the eigenvectors of the operator S :

$$S = \int_0^\tau R^*(0,t)R(0,t)dt. \quad (4.5)$$

Here τ is the forecast interval of interest, set at 24 months in this study, $R(0, t)$ is the forward tangent propagator of the TLM that advances the state vector of the system from time 0 to time t , $R^*(0,t)$ is the transpose of $R(0, t)$. A detailed description of the SO can be found in Moore et al. (2006), Tang et al. (2005), and Tang et al. (2008). Specifically, at each initial time, the perturbation was held constant for a total of 30 days, as a continuous wind perturbation following Karspeck et al. (2006), and then a new temporally uncorrelated perturbation was applied. The perturbations were controlled by (4.4) but using SO1_wind instead of SV1_sst, where X is still a normalized random number; and $\alpha = 0.7$ equivalent to the RMSE of winds anomaly of 0.7 m/s, obtained using sensitivity experiments based on the first verification principle Eq. (4.6) described in section 4.4.

4.3.4 Combination of Stochastic Optimal and Initial SSTA Perturbations

The fourth ensemble construction method is denoted by SO1_wind+SV1_sst, including the SV1_sst perturbation at initial conditions and the SO1_wind during the whole forecast period. Thus, two key sources of uncertainties were included in the SO1_wind+SV1_sst method. Comparisons between the SO1_wind+SV1_sst method against the SV1_sst method and the SO1_wind method, reflect relative importance of the uncertainty from the SO1_wind and from the SV1_sst in ensemble probabilistic predictions.

4.4. Verification Principles of Probabilistic Forecasts

ENSO probabilistic forecasts are made for three categories in this study: La Niña, Neutral, and El Niño. The category classification follows the definition⁺ used by the IRI ENSO forecast system, where the LDEO5 model is one of the forecast models used routinely for ENSO probabilistic forecast. Specifically, three ENSO categories are defined by the observed NINO3.4 SSTA binned at its climatological frequency of 25%, 50%, and 25%, respectively, which approximately match the common historical ENSO events during 1950-2002.

It is necessary to mention key properties of a probabilistic system here. A probabilistic forecast system has two key attributes: (i) Reliability, defined by statistical consistency between forecast probability (P_f , the proportion of ensemble members that indicate the occurrence of an event) and the corresponding observed frequencies (P_o) over the long time period (Toth et al. 2003). For example, the forecast system for precipitation is reliable if the proportion of occurrences of rain is close to P_o . However, reliability alone is not sufficient for a probabilistic forecast system. For example, a system always forecasting the climatological probability (P_c) of the event is reliable but not useful because the system would not provide any forecast information beyond climatology. Thus another key property of a probabilistic system is also required: (ii) Resolution, measures the difference between observed frequencies P_o and climatological probability P_c (Murphy, 1973). Compared to the P_c , a larger P_o indicates a higher resolution of the forecast system. Note P_o is obtained by compiling a set of cases for rain forecasts with P_f , P_o depends on P_f implicitly.

⁺<http://portal.iri.columbia.edu/portal/server.pt?open=512&objID=945&PageID=0&cached=true&mode=2&userID=2>

To achieve a reliable and high resolution ensemble-based probabilistic ENSO forecast, several principles used to measure the two properties are applied to evaluate our ensemble construction methods as discussed below.

4.4.1. Ensemble Spread and Error of Ensemble Mean

If the observation is statistically indistinguishable from the ensemble members, then the error of the ensemble mean (i.e., $RMSE_{EM}$) must close to the mean distance of the individual members from their mean (i.e. ensemble standard deviation or SPRD) (Buizza 1997; Stephenson and Doblas-Reyes 2000; Toth et al. 2003). In addition, the $RMSE_{EM}$ is comparable to the RMSE of the deterministic forecast ($RMSE_{CTL}$), obtained from the unperturbed initial condition. However, when nonlinearity becomes pronounced with increased lead time, the ensemble prediction could be better than the control forecast (Toth and Kalnay 1997). Furthermore, the standard deviation of the observed SSTA distribution over a long time period indicates the upper limit of RMSE for ENSO climatological predictions. With the observed NINO3.4 SSTA index for the period of 1856-2003, the standard deviation value is 0.71.

Thus, if an ensemble forecast system has good resolution, over a long time period, the following relationship is valid:

$$SPRD \approx RMSE_{EM} \leq RMSE_{CTL} \leq 0.71 \quad (4.6)$$

Note the SPRD in Eq. (4.6) is a function of lead time (t) only.

4.4.2. Reliability diagram (RD)

The traditional reliability diagram (RD; Wilks 1995) is often used to evaluate the reliability of probability forecast with two categories: the observed relative frequency of event occurrence (P_o) and the forecast probabilities (P_f). The P_o is calculated at a set of forecast probabilities from 0-100% in 10% intervals. The reliability diagram is a plot of P_o against P_f . If the forecast is perfectly reliable, the P_o should be similar to P_f . The RD method is good at evaluating and calibrating the reliability of a two-category (Yes/No) forecast. It also can be applied for a multicategory forecast by examining the reliability of individual categories separately. Also, one can evaluate the reliability by another method, the multicategory reliability diagram (MCRD) method (Hamill 1997).

4.4.3. The Brier Score

The Brier score (BS; e.g., Wilks 1995) is a commonly used verification measure for assessing the accuracy of probability forecasts. It is the mean squared error between the forecast probability and the observed frequency over the verification period.

$$BS = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \quad (4.7)$$

Where N is the number of total verification samples ($N=1776$ here), P_i is the forecast probability and O_i is a value 1 or 0 depending on whether the event occurred or not. Similar to the deterministic prediction skill RMSE, a smaller BS indicates a good forecast system.

The BS can be decomposed into three items: reliability (REL), resolution (RES), and

uncertainty (UNC) as follows (e.g. Wilks 1995)

$$BS = \underbrace{\left[\frac{1}{N} \sum_{k=1}^{K=10} n_k (P_{fk} - P_{ok})^2 \right]}_{\text{REL}} - \underbrace{\left[\frac{1}{N} \sum_{k=1}^{K=10} n_k (P_{ok} - s)^2 \right]}_{\text{RES}} + \underbrace{[s(1-s)]}_{\text{UNC}} \quad (4.8)$$

Over the verification period, the observed frequency of occurrence P_o can be partitioned into K bins ($K=10$ in this study) according to the forecast P_f . P_{fk} is the averaged forecast probability at bin k . P_{ok} is the corresponding observed frequency. s is the climatological probability (the base rate) that is independent of the forecast system. The uncertainty term UNC and base rate s are obtained from the long term observed data. For the cold, neutral, and warm ENSO category, UNC equals to 3/16, 4/16, 3/16 respectively according to the IRI definition mentioned earlier. n_k is the number of the forecast and observation pairs located in an individual bin k . The first term reliability RES on the rhs of Eq. (4.8) is actually equal to the mean squared deviation of the reliability curve from the diagonal line in RD plot. A smaller reliability term REL indicates a better consistency between P_k and \bar{O}_k , which results in a smaller BS and a more reliable probabilistic prediction skill. The second term resolution RES is equivalent to the variance of observed distribution. RES measures the ability of a forecast system to discern situations where the frequency of the occurrence of the event is different from the base rate s . Note that the RES term has a negative sign, but it is often used without the negative sign, as a positive-oriented measure.

A good Brier score occurs at a large RES item and a small REL, corresponding a high resolution and good reliability. The ideally perfect RES value equals to the uncertainty item UNC that gives the upper limit of the predictability of the probabilistic prediction system.

In order to compare the Brier score to that for a reference forecast system BS_{ref} , it is convenient to define the Brier skill score (BSS; Wilks 1995).

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (4.9)$$

If the climatological forecast is taken as reference prediction, $BS_{ref} = UNC = s(1-s)$. BSS is positively oriented. It has the range of $-\infty$ to 1. A negative BSS indicates that the forecast is less accurate than the climatology forecast. BSS equals to one for a perfect system, and zero for a system that performs like the climatology forecast.

From Eq. (4.8), Eq. (4.9) can be rewritten as

$$BSS = \frac{RES}{UNC} - \frac{REL}{UNC} = B_{res} - B_{rel} \quad (4.10)$$

In Eq. (4.10), B_{rel} and B_{res} are named as the reliability term and resolution term of the BSS score. B_{rel} is negatively oriented and B_{res} is positively oriented, that is in consistent with the signs of the RES and REL in the BS score. $B_{res} = 1$ and $B_{rel} = 0$ indicate a perfect forecast system.

4.4.4 The RPS Score

The ranked probability score (RPS; Epstein 1969; Murphy 1969; Murphy 1971) is another commonly used skill (resolution) measure for probabilistic forecasts, defined in terms of the squared differences between the cumulative probabilities in the forecast and observation vectors.

$$RPS(t, i) = \sum_{l=1}^3 \left(\sum_{k=1}^L p_k(t, i) - \sum_{k=1}^L O_k(t, i) \right)^2 \quad (4.11)$$

where P_k is the forecast probability assigned to the l th category and $O_k=1$ when the observation falls into l th category and 0 otherwise. The ranked probability skill score (RPSS; Wilks 1995) is defined using the RPS and a reference forecast defined to have zero skill. Here, the climatological forecast is used as the reference forecast.

$$RPSS = 1 - \frac{RPS}{RPS_{c\lim}} \quad (4.12)$$

where $RPS_{c\lim}$ is the RPS of climatological forecast. RPS/RPSS scores are functions of lead time (t) and initial time (i).

In summary, a good ensemble-based probabilistic forecast system should have: (i) Ensemble spread (SPRD) close to the $RMSE_{EM}$, and the RMSE of the control deterministic forecast, as given in Eq. (4.6); (ii) Probabilistic forecasts must be reliable, as measured by the reliability diagram and the reliability term of the Brier skill score (i.e. B_{rel}); (iii) A skillful probabilistic forecast system should have good resolution measured by the resolution term of the BSS score, (i.e. B_{res}). In addition, a skillful probabilistic forecast should have a small RPS, a large BSS/RPSS score.

4.5. Results

4.5.1 Ensemble Spread

We begin by first examining whether ensemble prediction experiments can satisfy the first principle Eq. (4.6). The SPRDs of four ensemble experiments are compared against the RMSE of the control run ($RMSE_{CTL}$) and the RMSE of ensemble mean ($RMSE_{EM}$) (Figs. 4.1 a-d). As discussed in section 4.1, the first principle offers a measure to judge that whether an ensemble construction can include sufficient uncertainties of the model. In Fig. 4.1a, although the $RMSE_{EM}$ for the SV1_sst method is close to the $RMSE_{CTL}$ and the standard deviation of the climatological forecast (0.71; the blue dash dot line), the ensemble SPRD underestimates the model uncertainty significantly. Of note is the decrease in ensemble SPRD at lead times of 10-17 months, suggesting a limitation of using linear SV theory in ensemble construction over long lead times. Thus, the SV1_sst method might not be a good ensemble construction strategy for the long-term ENSO ensemble predictions. Note that ensemble SPRD depends on the amplitude of random numbers applied on the SV1 pattern. Here α is set to 0.25 in Eq. (4.4) to represent realistic uncertainty of the initial SSTA (0.25 °C; e.g., Karspeck et al. 2006). Certainly, a large SPRD can be obtained by increasing α , but that results in a larger SPRD than the $RMSE_{CTL}$ and $RMSE_{EM}$ at the lead time of 6 months (Fig. 4.2), still violating Eq. (4.6).

In the second ensemble construction method (UV_realstoc), the high-frequency realistic stochastic winds are used during the forecast period. The current LDEO5 model is free of atmospheric random forcing, thus using realistic stochastic winds could potentially improve ENSO predictability. Unfortunately, the UV_realstoc method also underestimates the model error/uncertainty, showing a small SPRD far away from the $RMSE_{CTL}$ and the standard

deviation of the climatological forecast in Fig. 4.1b. Increase in the magnitude of external forcing can produce a large spread as shown in Fig. 4.3, however, artificial adjustment of the strength of stochastic winds results in unrealistic stochastic forcing. For example, the spread is close to the $RMSE_{CTL}$ when the perturbation magnitude is increased to three times of the original NCEP winds in Fig. 4.3. This is in agreement with the result from the LDOE4 model in Karspeck et al. (2006), where a sufficient spread could not be obtained until using an unrealistic strong wind forcing, with a standard deviation of 10 m/s. Fig. 4.4 shows that if the stochastic winds are unrealistically large (e.g. a strong wind perturbation 3.0 times as large as the original NCEP winds), the anomaly correlation skill R and RMSE degrade in spite of a large SPRD. An unrealistic strong wind perturbation may bias the model system and produce a large dynamical imbalance. Thus, the second strategy fails to construct a good ensemble forecast either.

In summary, both the SV1_sst and the UV_realstoc methods cannot introduce sufficient uncertainties that we expect for a good ensemble construction. For the SV1_sst method, large differences between the SPRD and $RMSE_{CTL}$ at longer lead times suggest that the perturbation introduced at the initial SSTA cannot effectively persist through the forecast period due to dispersion. For the UV_realstoc method, uncertainty estimated from the high frequent components of NCEP winds cannot produce sufficient prediction uncertainties or errors due to the random nature of the perturbation spatial structure. As mentioned in section 4.3.3 and in the introduction section, the spatial structure of stochastic wind perturbation is important in ensemble construction.

In the third experiment, we used the stochastic optimal mode to construct the ensemble prediction for the period from 1856-2003, as discussed in Kleeman and Moore (1997) and in

section 4.3.3. To achieve this, we first calculated the leading SO mode of winds (denoted by SO1_wind) for each calendar month for the optimal period of 24 months over the 148-yr period. It was found that the spatial pattern of the SO1_wind is not sensitive to initial conditions, thus the average SO1 wind pattern over all initial conditions, as shown in Fig. 4.5, was used for the ensemble construction. Similar to the SV1 of SSTA, the SO1_wind contributes to about 30-40% of the total variance. As seen in Fig. 4.5, there is a strong convergence region of winds centered at 140 °W and a divergence at the cold tongue region of the eastern tropical Pacific. That such a structure is favorable for perturbation growth is probably inherent in ENSO dynamics. For example, this pattern generates corresponding downwelling and upwelling in the eastern tropical Pacific, and induces warm eastward-propagating Kelvin waves and cold westward-propagating Rossby waves, which in turn impacts on ENSO variability according to the delayed oscillator theory (Suarez and Schopf 1988). Fig. 4.1c shows the SPRD variation as a function of lead time, generated by the SO1_wind method. As can be seen, the $RMSE_{EM}$, SPRD from this method are closer to the $RMSE_{CTL}$ and the standard deviation of ENSO climatological prediction (0.71) than the first two methods, satisfying the first principle Eq. (4.6). Comparison of Fig. 4.1b with Fig. 4.1c suggests the importance of the spatial structure of perturbation in ensemble construction. Note that the perturbation magnitude used here is much smaller than that of UV_realstoc (0.7 m/s against 2.5m/s).

The fourth perturbation ensemble construction method (SO1_wind+SV1_sst) is to combine the SV1_sst and the SO1_wind perturbations. In terms of the first principle Eq. (4.6), the ensemble spread produced by this method is the best, as shown in Fig. 4.1d. Compared with the SO1_wind and the SV1_sst method, the SPRD from SO1_wind+SV1_sst method is the

closest to the $RMSE_{EM}$ and $RMSE_{CTL}$, showing the important effect of both the $SV1_{SST}$ and the $SO1_{wind}$ perturbation on the ensemble spread. Especially, $SV1_{SST}$ perturbation mainly impacts the spread of short-leading times whereas $SO1_{wind}$ likely dominates the ensemble spread of long-leading times.

4.5.2 Reliability

The second principle, “reliability”, is examined by the reliability diagram (RD) method. The forecasted/observed SSTA are grouped into three categories representing the cold, neutral, and warm ENSO states, as defined at the beginning of section 4.4. In each ENSO category, a RD curve is made by using the forecast probability P_f at 11 intervals of 0, 10%, ..., 100% against the corresponding observed relative frequency P_o over the 148 years. The diagonal line in a RD diagram indicates a perfect reliable system, i.e. $P_f = P_o$.

RD diagrams are shown in Fig. 4.6 for the four ensemble construction methods and at three different lead times: 6, 9, and 12 months. The RD curves from the first two ensemble construction methods cross the diagonal line from the upper-left to bottom-right showing poor reliability (Figs. 4.6 a-b). These features are probably due to the low “resolution” of the first two methods. Note that the resolution of probabilistic forecasts can also be approximately estimated in the reliability diagram. According to the definition of resolution in Section 4.4, the resolution is determined by the difference between P_o and P_c . When a RD curve becomes flattened, it is closer to its climatological probability values (that is, P_o tends to be close to the P_c , 0.25/0.5/0.25 for the cold/neutral/warm category, respectively). For the last two ensemble construction methods, in Figs. 4.5 c-d, reliability is greatly improved, since the RD curves oscillate around the diagonal lines, especially for the fourth method $SO1_{wind}+SV1_{sst}$.

Reliability can be quantified using the reliability component of the Brier skill score (B_{rel} ; Wilks 1995). Fig. 4.7 shows the reliability scores for four ensemble experiments for the three ENSO categories. Again, the two SO-based ensemble construction methods provide more reliable results (i.e. smaller reliability scores) than the other two methods over all lead times and in all categories. Comparing the reliability scores of SO1_wind with the SO1_wind+SV1_sst, the later method is superior, due to the contribution of the SV1_sst at short lead times. Thus, both the RD analysis and the reliability score demonstrate that the importance of the stochastic optimal winds in the ensemble construction.

To illustrate the role of ensemble SPRD on reliability, it is more convenient to use the verification rank histogram. A rank histogram diagram is another way to present the reliability of an ensemble forecast system. The underlying reason is that the observation and ensemble members in a reliable ensemble system are subject to an identical probability distribution. In a rank histogram, each of the $M+1$ intervals, defined by an ordered M ensemble members, includes two open ended intervals. A reliability system would be equally likely to contain the observed value (Toth et al. 2003). In a small SPRD case as seen in Fig. 4.7, the rank histogram plot displays “U” type distribution. Observations fall more frequently on the first and the last categories and rarely show in the middle categories. This is because for a small SPRD case, the observation will often be an outlier in the distribution of ensemble members, implying that ensemble relative frequency will be a poor approximation to the probability. In a good SPRD case, as seen in Fig. 4.8, the rank histogram shows a homogenous distribution, i.e. the frequency distribution is around the perfect percentage line, indicating the consistency of forecast and observed distributions (a good reliability). Thus, an ensemble system with good ensemble SPRD will result in a reliable probabilistic forecast system.

4.5.3 Resolution

To examine the resolution of the four ensemble construction methods, Fig. 4.10 displays the resolution terms of the BSS score (B_{res}) for the warm, cold, and neutral ENSO states, as a function of lead time. Two common features can be seen: (i) The B_{res} scores for the warm and cold ENSO events are greater than those of the neutral ENSO state for a given lead time, and resolution drops faster at the neutral ENSO state than the others, indicating that El Niño and La Niña events are more predictable than neutral events. This signal-dependent characteristic of ENSO predictability is in agreement with many studies (e.g., Chen and Cane 2008; Tang et al. 2008a). (ii) Compared with the large differences of reliability terms among four methods in Fig. 4.7, resolution terms for four methods only show slight difference, although their SPRD are visibly different in Fig. 4.1. This implies that SPRD is more related to reliability than resolution. In other words, the reliability of ENSO probabilistic forecast is more sensitive to choice of ensemble construction strategy than does the resolution.

According to the definition of resolution, a more skillful probabilistic forecast system requires a larger difference between the observed distribution and the climatological distribution. Because observations are arranged in different bins based on the forecast probability, resolution measures ability of a forecast system to discern types of events. It seems there is no explicit relationship between the ensemble SPRD and the resolution. For example, in an extreme case, if ensemble SPRD and ensemble mean are invariant over the verification period, the ensemble system will have a poor “resolution” no matter how large (or small) the SPRD is. In this specific case, because the system provides a constant forecast probability in each ENSO category, forecasts and observations will be aggregated in one bin, the observed

frequency will be equal to the base rate, and the resolution will be zero. Under this consideration, the temporal variability of ensemble spread, rather than the SPRD itself, may play an important role in determining the resolution.

4.5.4 Overall Probabilistic Skill

The overall performance of the four ensemble construction methods is evaluated by BSS score and RPS/RPSS score. The BSS score includes the joint contribution of reliability and resolution at each ENSO category. The RPS and RPSS are accumulated skill scores for all categories. Fig. 4.11 presents BSS for four ensemble methods at cold, neutral, and warm ENSO states. As seen, the SO1_wind and SO1_wind+SV1_sst methods provide better skill scores than the SV1_sst and UV_realstoc method, indicating the important role of stochastic optimal winds in improving probabilistic skill. The larger BSS scores in the last two methods mainly benefit from the better reliability terms in Fig. 4.7, because four experiments have the similar resolution terms as seen in Fig. 4.10. In addition, the SO1_wind+SV1_sst method shows a larger BSS score than the SO1_wind method at short lead time, due to the smaller reliability term of the SV1_sst method at short lead time. Fig. 4.11 also indicates the upper limit of ENSO predictability of the LDEO5 model using BSS score. Warm and cold ENSO events are predictable for lead times of 2 years or longer, whereas the neutral ENSO state reaches its lowest predictability at the lead time of 10 months (i.e., at the lead time longer than 10 months, the BSS is negative, indicating the system has no skill at longer lead times).

The RPS and RPSS scores measure the distance between the probability of the forecast and observation similar to the RMSE value, but in probabilistic sense. From the definition of RPS, the range of RPS score is between 0 (the perfect forecast) and 1. The RPSS score is zero or

positive if the forecast skill equal to or exceeds that of the climatological probabilities, while a negative RPSS represents that the forecast skill is worse than climatology (e.g., Mason 2004). A smaller RPS or a larger RPSS score indicates higher predictability. To compare the skill of the four ensemble methods, individual RPS and RPSS scores were calculated over 148 years for lead times from 0 to 24 months. The averaged RPS and RPSS score over the 148 years are given in Fig. 4.12. The SO1_wind+SV1_sst method has the smallest RPS score and largest RPSS score: it provides a more skillful forecast than the other methods. It is worth noting that the RPSS scores shown in Fig. 4.12 are averaged over three ENSO categories for lead times of 0-24 month over the 148 years, thus, although the averaged RPSS scores have negative values at lead times >5 months, for individual forecasts of warm or cool events the skill score can have positive RPSS scores.

In summary, Fig. 4.11 and Fig. 4.12 indicate that the fourth ensemble construction method SO1_wind+SV1_sst is superior to the other three, providing the highest probabilistic prediction skill. Also, the third method SO1_wind has relatively higher prediction skill than the first and the second methods. Thus, we have demonstrated that the stochastic optimal winds play important roles in constructing ensemble prediction in the LDOE5 model.

4.6. Conclusion and discussion

Skillful ENSO predictions will assist in the management of natural resources and the environment. Significant progress has been made in ENSO prediction over the past few decades (e.g., Latif et al. 1998; Shukla et. al. 2000; Goddard et. al. 2001). Currently there are a few ENSO prediction models issuing routine predictions (e.g., IRI at <http://portal.iri.columbia.edu>), including statistical models, intermediate complexity dynamical

models, hybrid coupled models, and fully coupled general circulation models. However, some important issues still remain and are challenging to the ENSO and seasonal climate prediction community. One specific issue is the measures of the uncertainties in ENSO prediction.

An ideal approach to deal with prediction uncertainty is to issue probabilistic prediction, which has been widely applied in weather forecasting. Compared with weather probabilistic forecasting, ENSO probabilistic prediction has not been well addressed. Probabilistic predictions are typically generated by ensemble prediction methods. Thus an interesting question is: which ensemble construction method can lead to the best ENSO probabilistic model? In this study, we explored four typical ensemble construction methods through the LDOE5 model. A long term retrospective ensemble prediction was carried out for the past 148 years (1856-2003) for each ensemble construction method. The performance of probabilistic prediction is measured using several probabilistic verification methods (the spread principle, ROC, RD and the reliability score, RPS, and RPSS). The reliability, resolution, and the amplitude of the ensemble spread were considered as the key principles to evaluate the performance of ensemble construction methods.

It was found that the SV1_sst ensemble construction method and the realistic stochastic winds method UV_realstoc failed to generate reliable and high resolution probabilistic predictions because they characterize insufficient uncertainties in the model, resulting in small ensemble spreads. Meanwhile, their lower reliability and poor resolution are revealed by several probabilistic methods including: reliability diagrams, reliability scores, ROC scores and RPS/RPSS scores. The small spreads in both of the SV1_sst and UV_realstoc methods are probably due to the limitation of linear SV theory at the longer lead times and due to the random nature of spatial structure in the high-frequency realistic winds, respectively. To

overcome the small spread issue, stochastic optimal perturbation of winds were applied over the whole forecast period in the last two SO-based methods. After removing the spread issue, the two SO-based methods exhibit good reliability in probabilistic measures.

In some specific cases, a good reliability system does not warrant skillful forecasts. For example, an unrealistic strong wind perturbation also can provide a good reliability, but it has a poor prediction skill because that an unrealistic wind perturbation could bias the model dynamics. Nevertheless, a good reliability prediction is useful to obtain high ENSO probabilistic skill (e.g. BSS and RPSS). Among four ensemble construction methods, BSS and RPSS indicate that the SO1_wind+SV1_sst ensemble construction method is superior to the other three. Also, the third method SO1_wind has a higher BSS score than the first and the second methods, indicating the stochastic optimal winds play important roles in constructing ensemble prediction in the LDOE5 model. The skillful perturbation method benefits from the good reliability at longer lead times contributed by the stochastic optimal winds and at shorter times by the SV1_SSTA. Furthermore, because the “resolution” is similar to each other in the four experiments and there are large differences in the reliability term, suggesting that ENSO probabilistic prediction skill more relies on the reliability term than on the resolution term.

One interesting finding in this study is the great importance of stochastic forcing on ENSO probabilistic prediction. Generally, there are two kinds of sources that limit ENSO predictability: the chaotic behavior of the nonlinear dynamics of the coupled system (e.g., Jin et al. 1994; Chen et al. 2004); and the stochastic nature of the coupled system characterized by weather noise and other high-frequency variations, such as westerly wind bursts and Madden-Julian Oscillation (e.g., Penland and Sardeshmukh 1995; Kleeman and Moore 1997; Moore et al 2006; Gebbie et al. 2007). It is still not clear which source plays the more dominant role.

Thus, the importance of stochastic forcing on ENSO probabilistic prediction provides insight on this central question challenging the ENSO community.

Several cautions should be borne in mind. First, we only investigated four ensemble construction methods. Recent studies of Ham et al. (2009), Zheng et al. (2009) suggested that the ENKF data assimilation approach is good ensemble construction method that can provide reliable and high resolution ensemble predictions. Thus further comparisons of the SO-based methods with other methods such as ENKF and ET methods are expected. Second, we only perturbed two variables (i.e. the SSTA and anomalous winds), other variables could also have important impacts on ENSO predictability. For example, Karspeck et al. (2006) suggested that thermocline depth (H1) or subsurface temperature (Te) has large impacts on error growth and predictability in the LDEO4 model. However, because SSTA is the only initial conditions used in the LDEO5 model, choosing the errors and uncertainties from SV1_sst at the initial time and using SO1_wind to represent external atmospheric wind noise seems to be a reasonable way of perturbing the LDEO5 model.

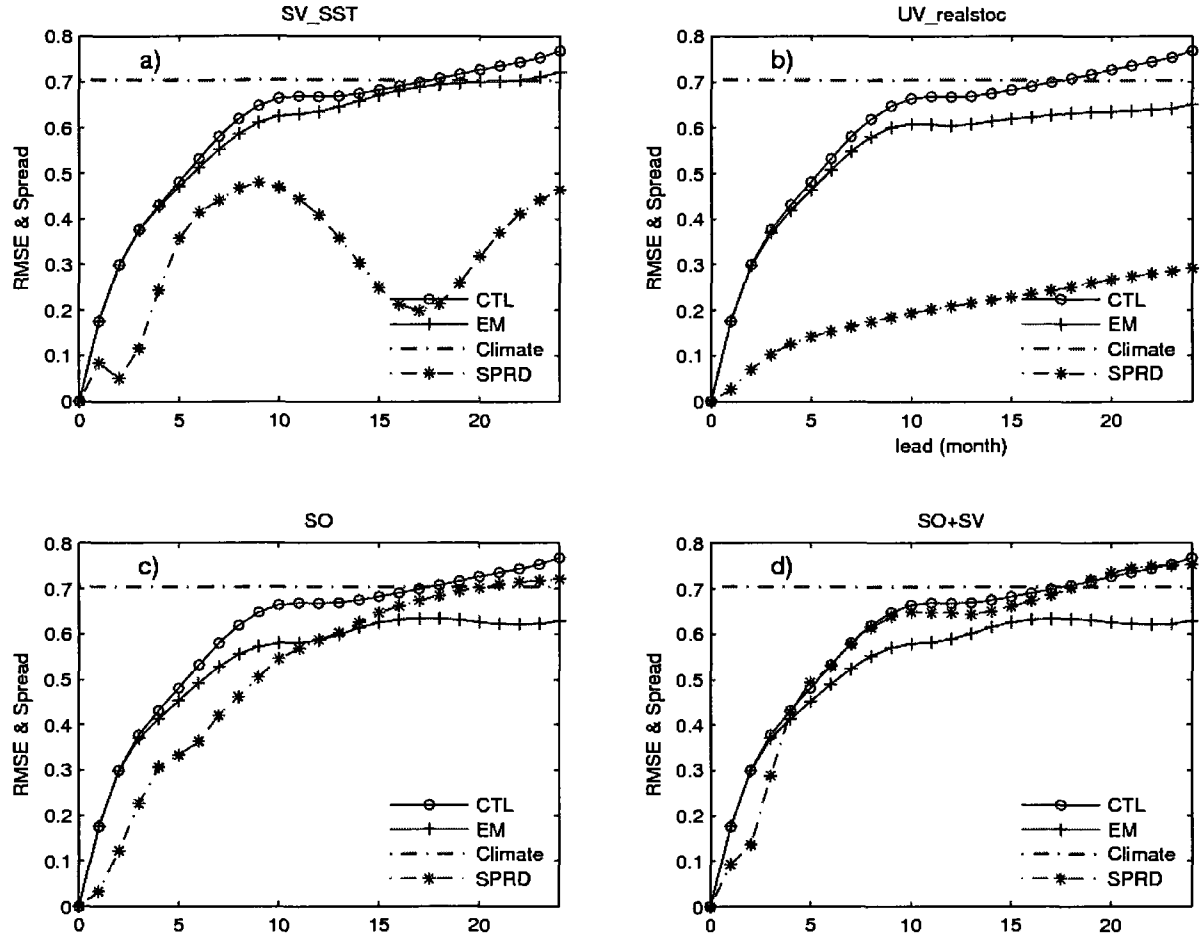


Fig. 4.1 Root mean square error (RMSE) of NINO3.4 SSTA for the control run (CTL, circle), ensemble mean ($RMSE_{EM}$, '+'), along with ensemble spread (SPRD, star), and climatological standard deviation of SSTA (dash-dot line) as a function of lead time (month), averaged over 1856-2003. a) RMSE & SPRD for SV1_sst method; b) UV_realstoc; c) SO1_wind method; d) SO1_wind+SV1_sst method.

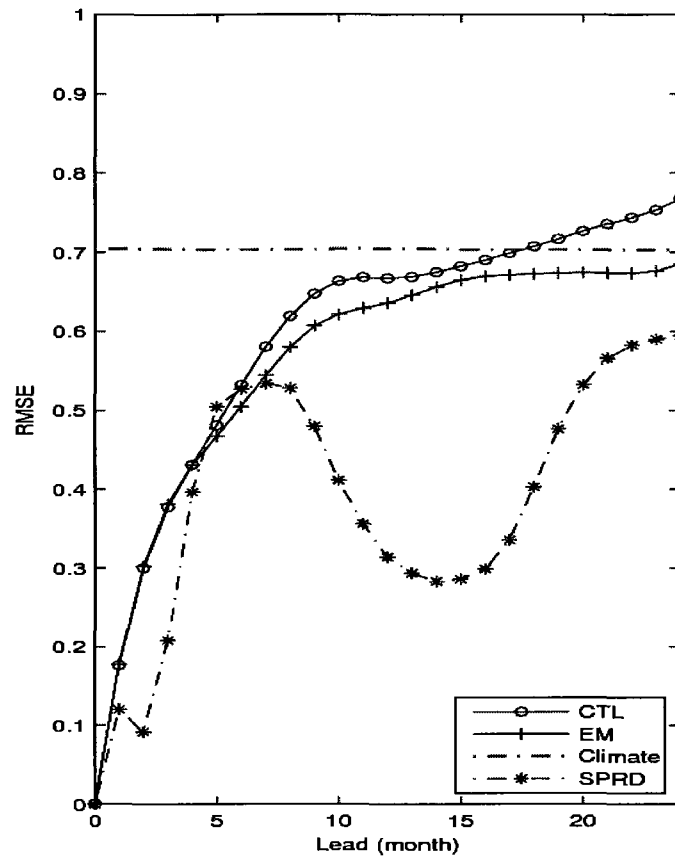


Fig. 4.2 Same as Fig. 4.1 but for the first ensemble construction method SV1_SST, with a larger SSTA initial perturbation magnitude (1.5 times of that in the Fig. 4.1 a)

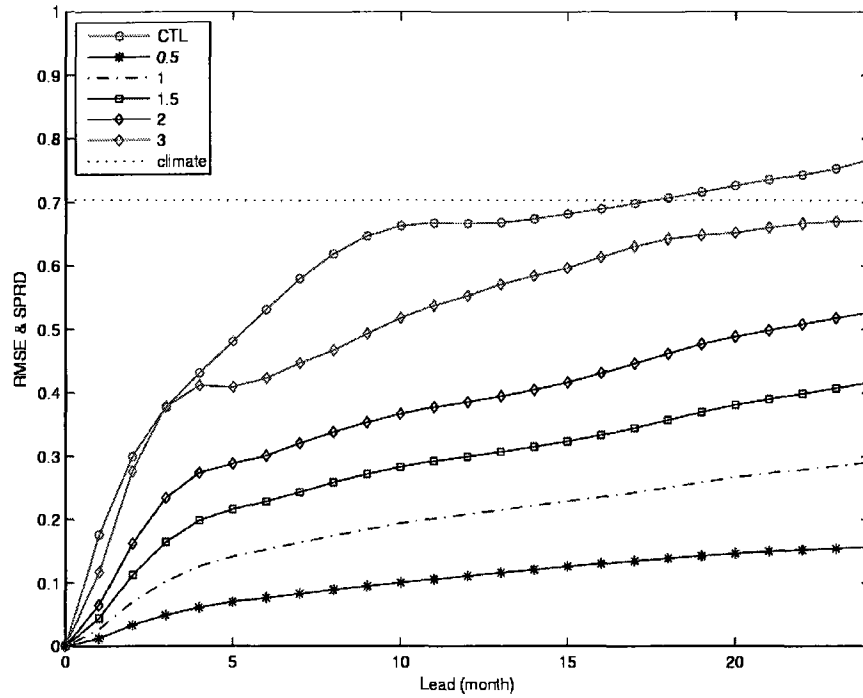


Fig. 4.3 A sensitivity study for the SPRD by adjusting the strength of stochastic winds in the second ensemble construction method. The perturbation magnitude varies from 0.5 to 3.0 times of that of NCEP high-frequency winds.

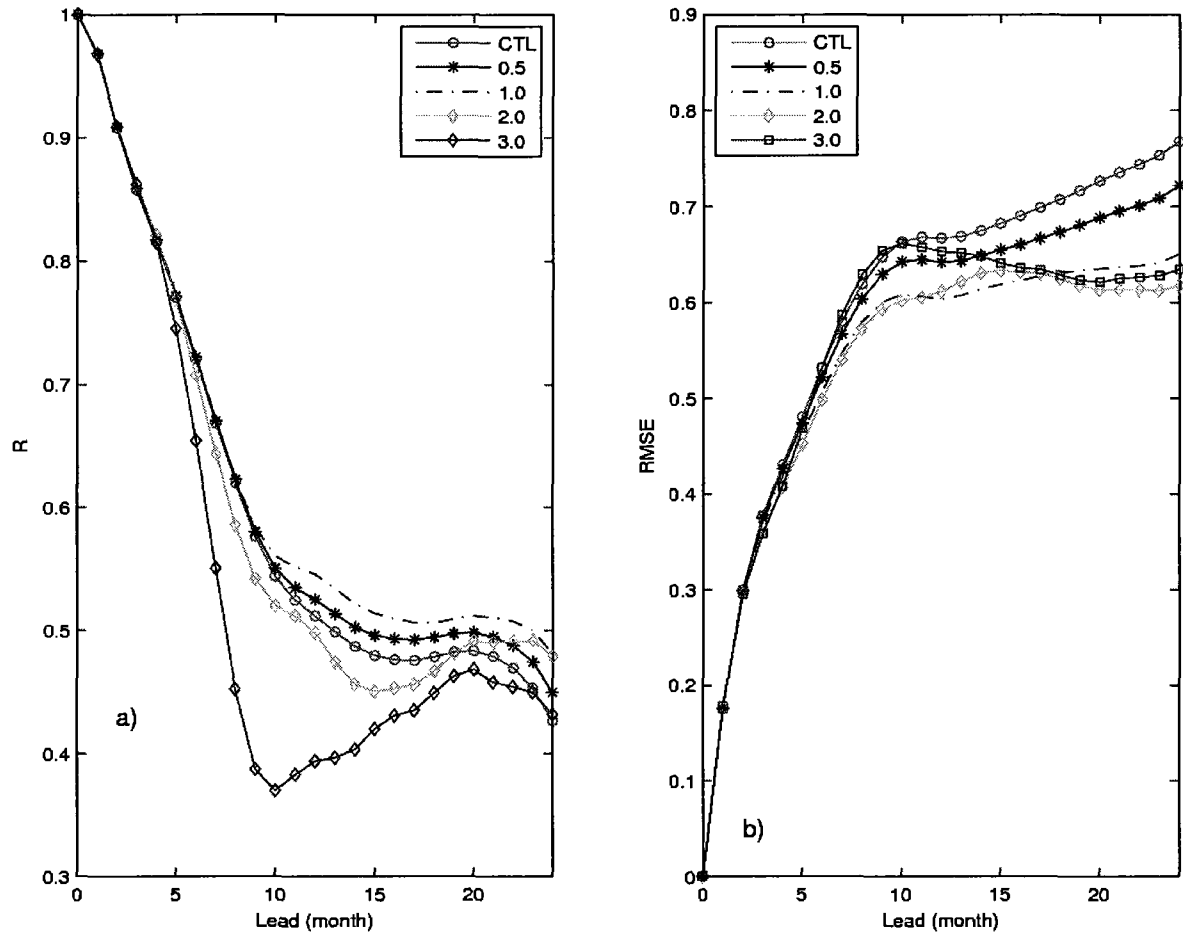


Fig. 4.4 a) Anomaly correlation skill (R) from the control run (open line), ensemble mean forecast with different level of stochastic wind perturbation (0.5, 1, 2, 3 times) than the realistic NCEP winds. b) Same as a), but for RMSE.

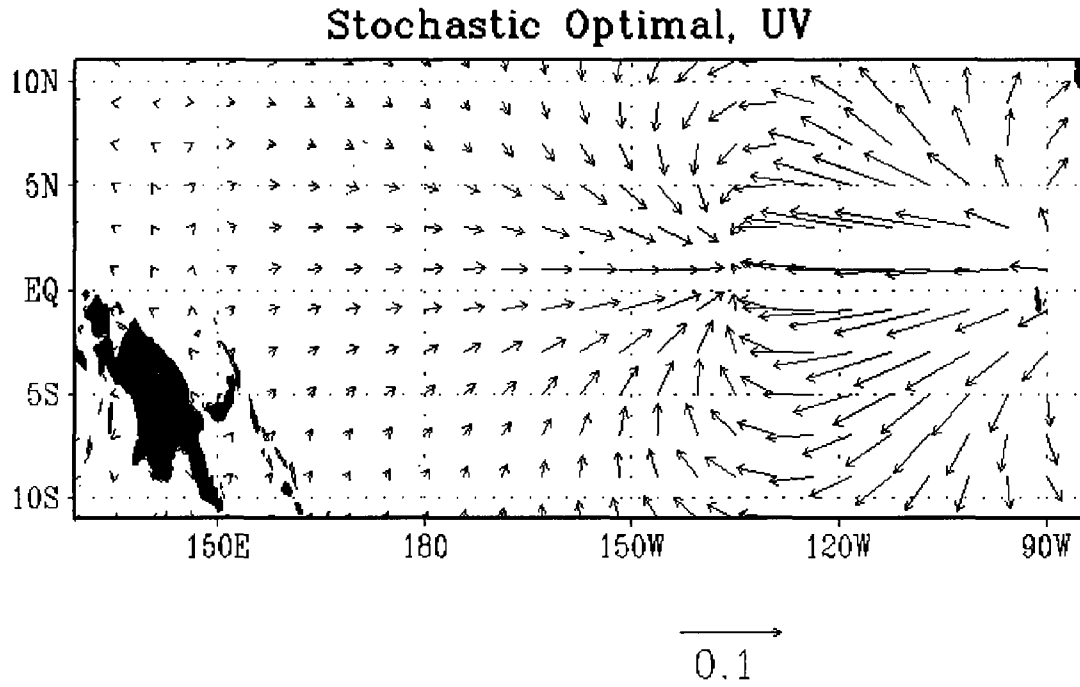


Fig. 4.5 The 148-yr averaged leading mode of the stochastic optimal (SO) winds (m/s). This mode explains the 30-40% of the original variance of S in Eq. (4.5).

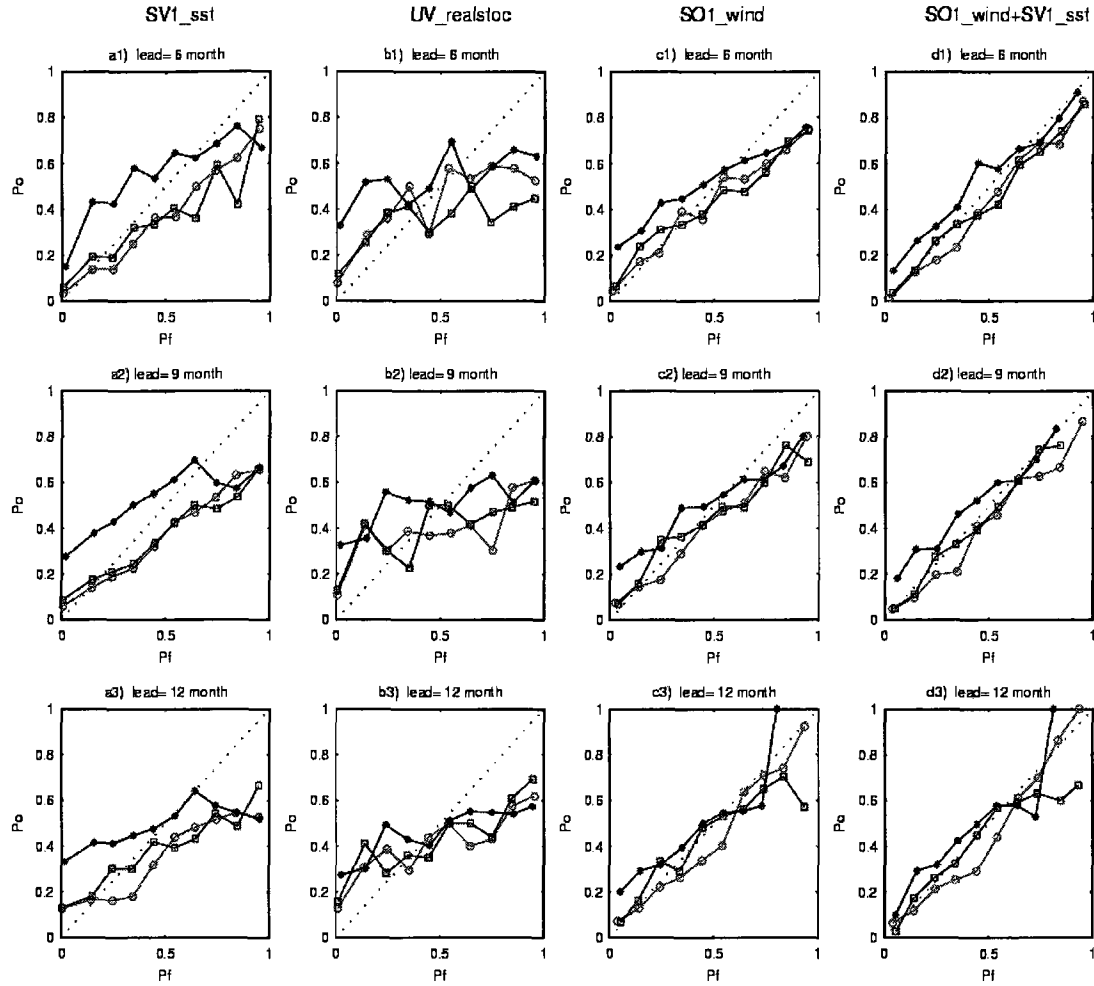


Fig. 4.6 a1-a3) Reliability diagram (RD) for the first ensemble construction method: SV1_sst, at lead times of 6, 9, 12 months. In each plot, three reliability curves represent three ENSO categories: Warm (circle), neutral (star), and cold events (square). These are calculated based on 100-member ensemble hindcasts for all months over the 1856-2003. b1-b3) for the second method: UV_realstoc; c1-c3) SO1_wind; d1-d3) SO1_wind+SV1_sst.

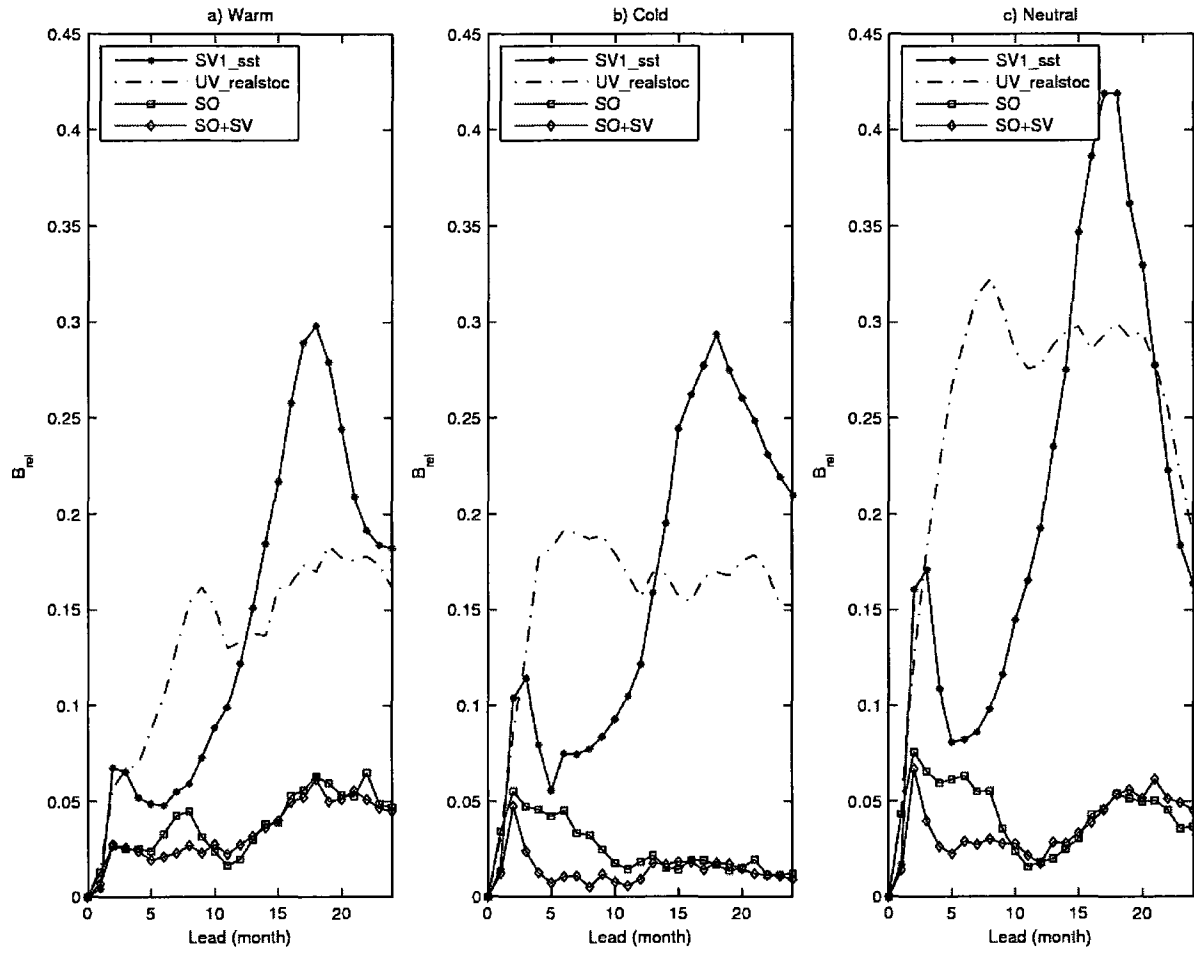


Fig. 4.7 The reliability term B_{rel} of the Brier skill score as a function of lead time for the four ensemble prediction experiments at a) warm, b) cold, and c) neutral ENSO categories as functions of lead time (month). (Star: SV1_sst; dash-dot: UV_realstoc; square: SO1_wind; diamond: SO1_wind+SV1_sst). B_{rel} is negatively oriented.

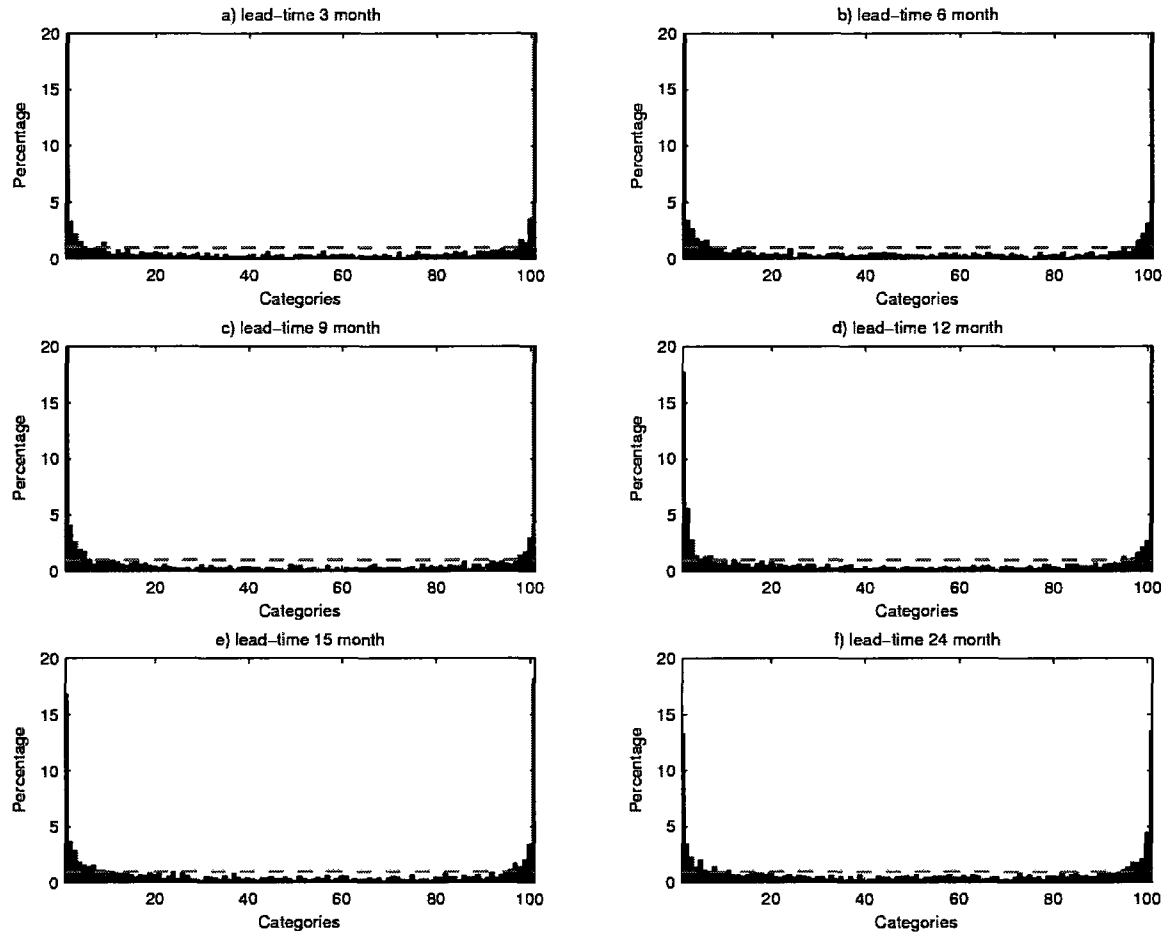


Fig. 4.8 Analysis rank histograms for a small SPRD case (i.e. the second ensemble construction method UV_realstoc using the original high frequency winds) at different lead times 3, 6, 9, 12, 15, and 24 months. In a 100 ensemble members, the perfect percentage is 1.0% (dash line).

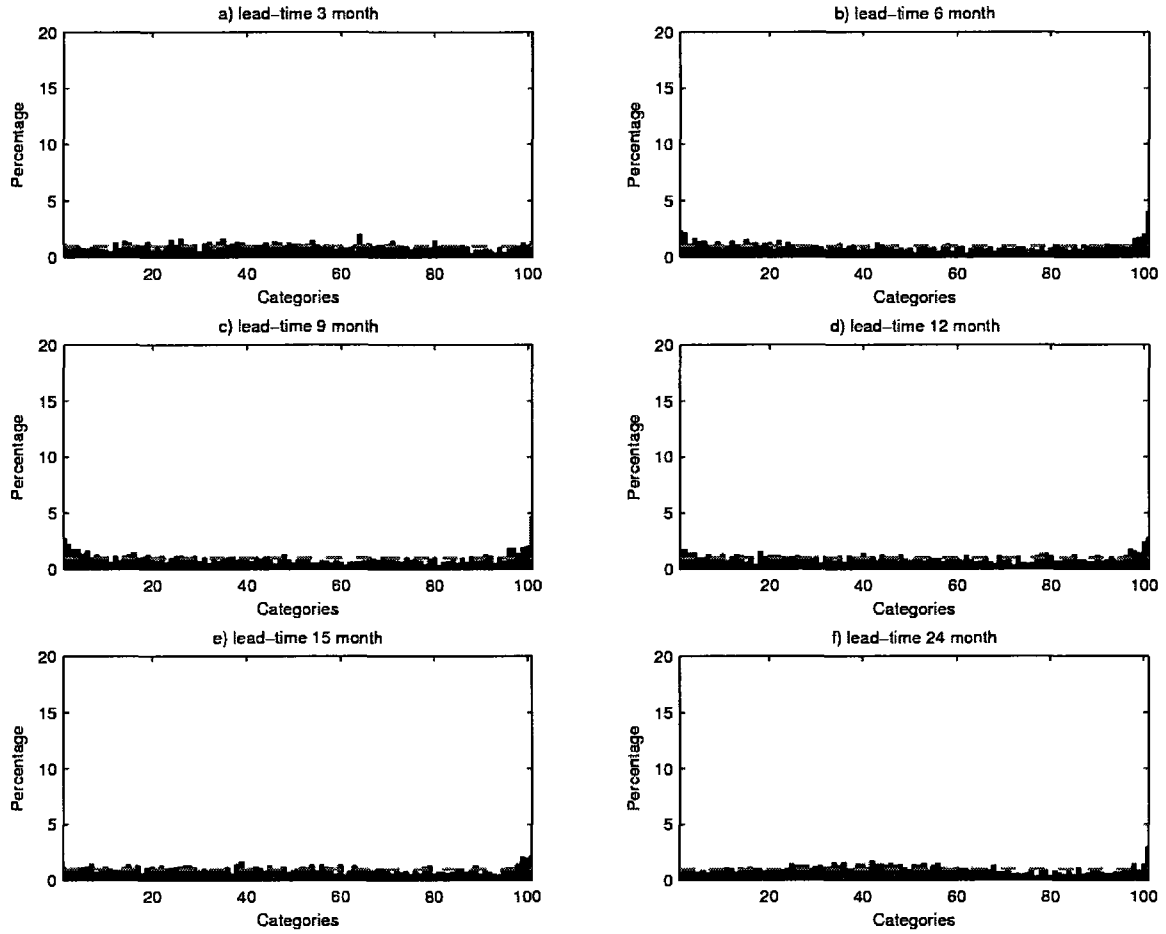


Fig. 4.9 Same as Fig.4.8 but for a good/sufficient SPRD case, with a strong wind perturbation (3 times of the original high-frequency NCEP winds).

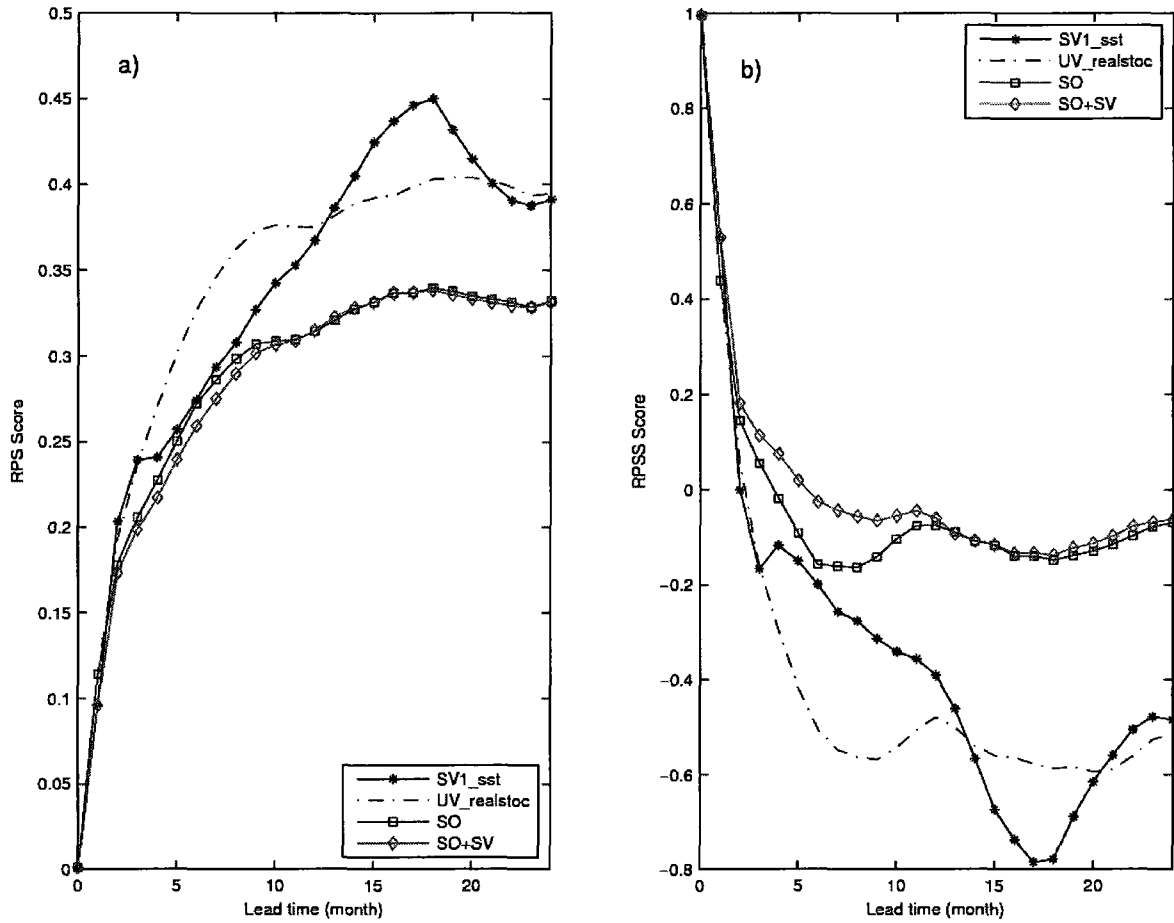


Fig. 4.10 a) The ranked probability score (RPS) as a function of lead time for three ensemble construction methods. b) Same as a) but for the ranked probability skill score (RPSS). (Star: SV1_sst; dash-dot: UV_realstoc; square: SO1_wind; diamond: SO1_wind+SV1_sst).

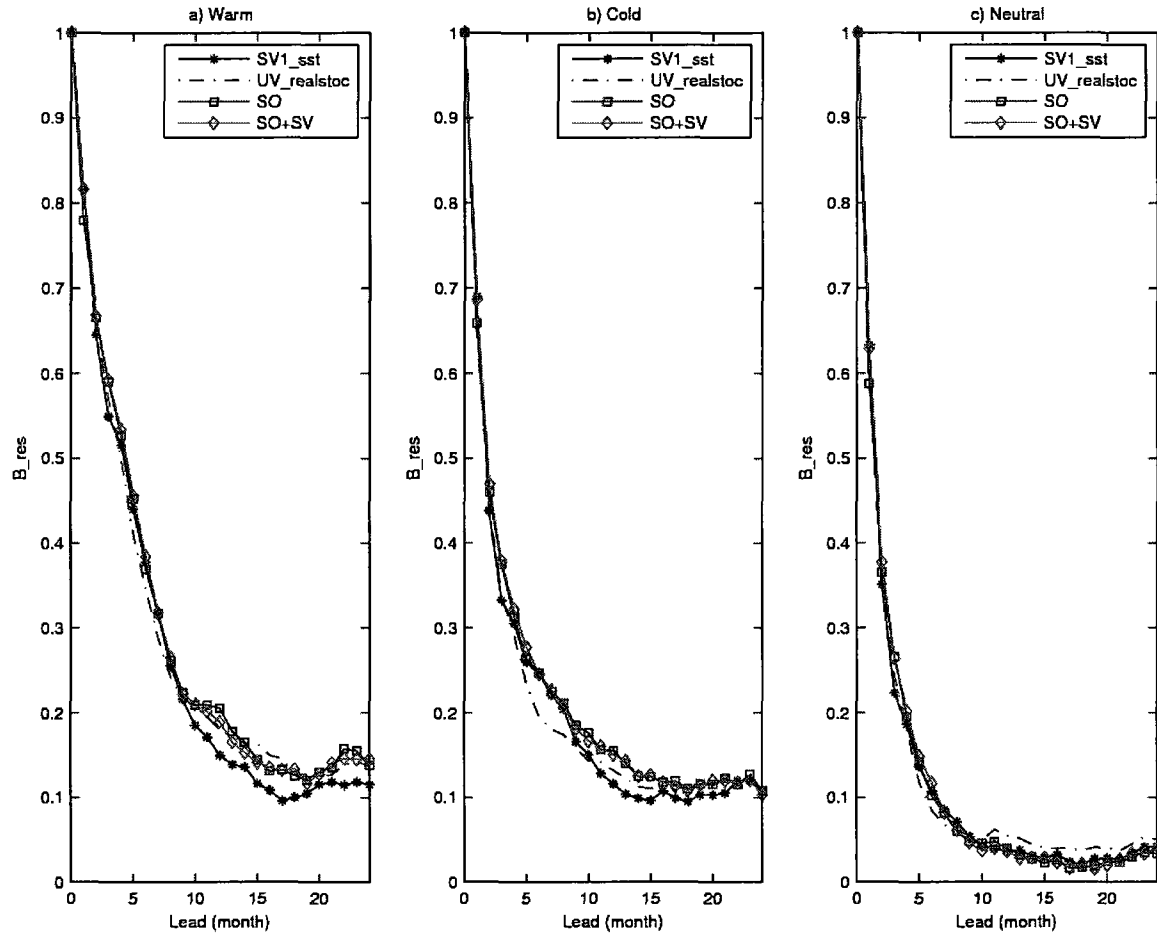


Fig. 4.11 Same as Fig. 4.7 but for the resolution term B_{res} of the Brier skill score. B_{res} is positively oriented..

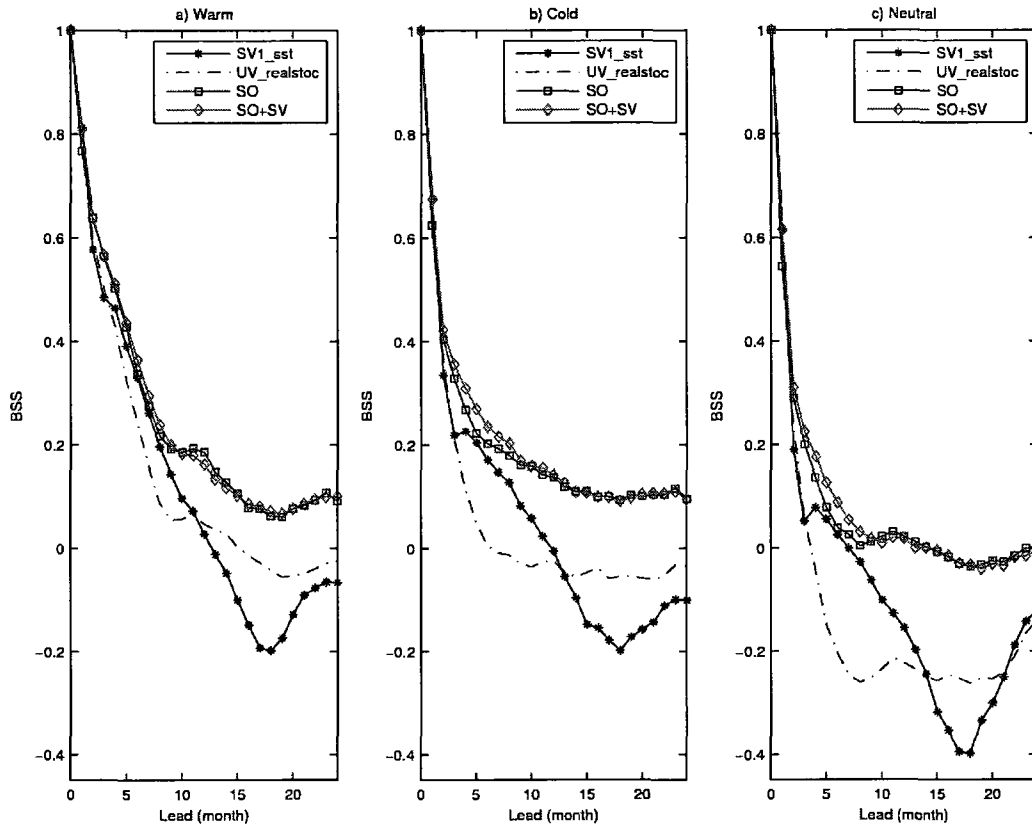


Fig.4.12 Same as Fig. 4.7 but for the Brier skill score (BSS). BSS is positively oriented.

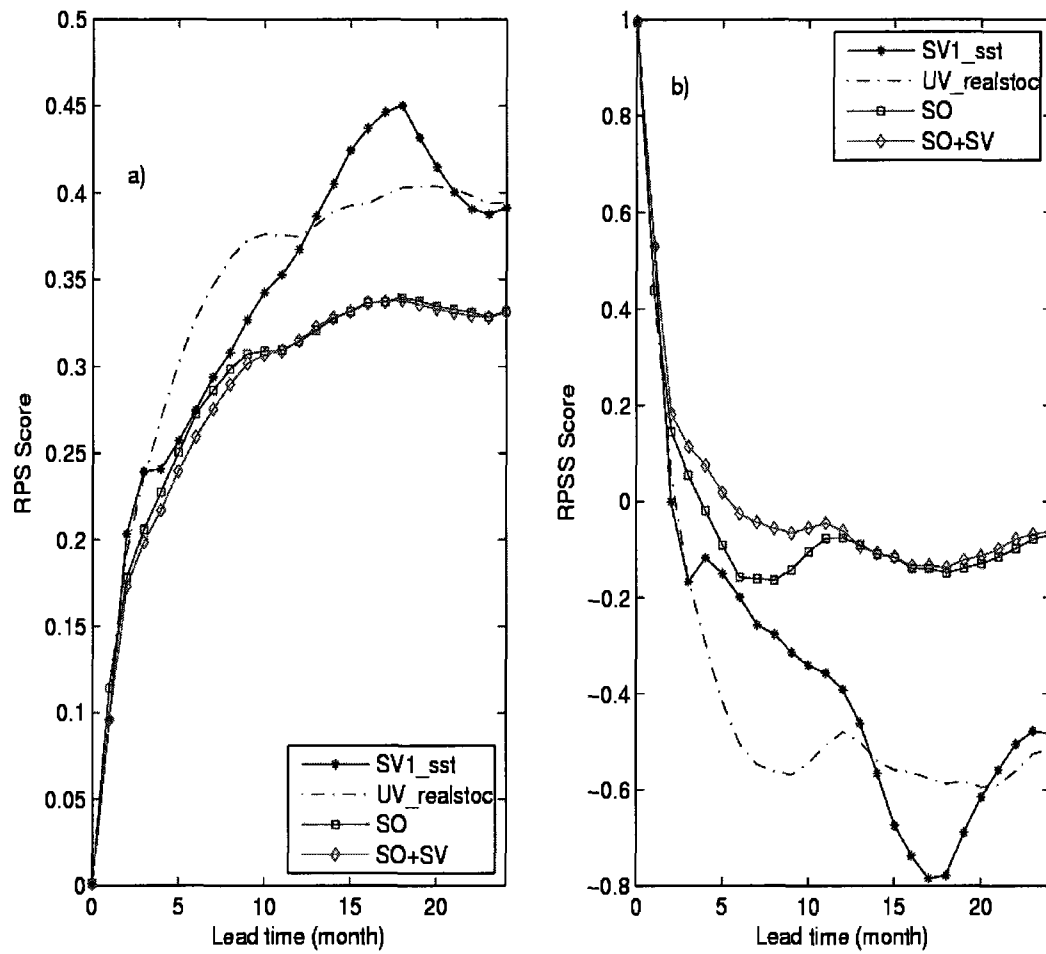


Fig. 4.13 a) The ranked probability score (RPS) as a function of lead time for four ensemble construction methods. b) Same as a) but for the ranked probability skill score (RPSS).

Chapter 5: ENSO Ensemble Prediction in the LDEO5 Model from 1856-2003: Potential Predictability vs Real Skill

Cheng Y, Tang Y, Chen D, Jackson P (2009) Ensemble Construction and Verification of the Probabilistic ENSO Prediction in the LDEO5 Model, (submitted to Climate Dynamics, in January 2009).

5.1. Introduction

Predictability is the study of the extent to which events can be predicted (e.g., DeSole 2004). Generally, there are two groups of predictability measures; one is actual measures that make use of observations, and the other one is potential predictability measures that do not make use of observations. For actual measures, e.g. the mean square error (MSE) skill indicates the mean square difference between forecasts and observations over the verification time period. MSE increases with lead time and asymptotically approaches a “saturation” value. The saturation value is equivalent to the mean square difference between two randomly chosen fields from the system (e.g., DeSole 2004).

An important task in ENSO predictability study is to quantitatively estimate predictability by potential predictability measures, by which the degree of confidence that can be placed in an individual forecast can be assessed (Moor and Kleeman 1998; Tang et al. 2008a). Traditionally, ensemble-based potential predictability measures include: ensemble mean (EM), ensemble spread (ES), and ensemble ratio (ER; the ratio of |EM| over ES) (e.g., Buizza and Palmer 1998; Moore and Kleeman 1998; Scherrer et al. 2004). A large ES generally suggests a relatively low predictability in ensemble weather forecast. However, these ensemble potential measures have often met with challenges and limitations in quantifying ENSO and climate prediction skill (e.g., Kumar et al. 2000; Tang et al. 2005; Tang et al. 2007; Tang et al. 2008a, b). Using 20-yr retrospective predictions of two hybrid ENSO models, Tang et al. (2008b) found that ensemble spread ES is not a good predictor in quantifying climate prediction skill in comparison with the

ensemble mean square (EM2).

Recently, new ideas from information theory have made their appearance to examine ENSO and seasonal climate predictability (e.g. DelSole and Tippett 2007). Several information-based potential measures have been used to qualify the potential predictability, such as relative entropy (RE), predictive information (PI), predictive power (PP), and mutual information (MI) (Schneider and Griffies 1999; Kleeman 2002, 2008; Tippett et al. 2004; Tang et al. 2005, 2008a; DelSole 2004, 2005; DelSole and Tippett 2007, 2008). Like the ensemble-based potential prediction skill metrics, information-based skill metrics also have the advantage that measures predictability without the use of observations.

Information-based measures have several important characteristics. First, for a normally distributed, stationary, Markov process, predictability declines monotonically with the length of the forecast due to chaos (Kleeman 2002; DelSole 2004; Tang et al. 2008a). Second, at a given lead time, the averaged RE and PI over a long time period should be identical, equal to MI (DelSole and Tippett 2007). MI indicates the overall prediction skill of a forecast system. Tang et al. (2008a) firstly examined these characteristics of information-based potential predictability measures for ENSO retrospective predictions using two realistic hybrid models. Their results show that the aforementioned characteristics of information-based measures generally held well in their ENSO models. However, their study only focused on 18 years from 1881-1998, so that the period available to test predictability covers rather limited ENSO cycles (typically 3-5), basically precluding statistically robust conclusions. In addition, it has been well recognized

that the actual predictability of ENSO has striking decadal/interdecadal variations (e.g., Chen et al. 2004; Tang et al. 2008). One might be able to shed light on the mechanism of decadal/interdecadal variation in ENSO predictability by exploring decadal/interdecadal variation of potential predictability. Obviously, the previous analysis for only an 18-yr period is unable to achieve this goal.

Chen et al. (2004) used KAPLAN sea surface temperature anomaly (SSTA) reanalysis data and the ZC model version LDEO5 (LDEO5 hereafter) to perform a 148 year hindcast experiment for the period of 1856-2003. They successfully predicted all of the prominent El Niños during this period at lead times of up to two years, with the SSTA being the only data used for model initialization. Tang et al (2008c) further analyzed the interdecadal variation in ENSO prediction skill from 1881-2000 using multiple models. These retrospective ENSO predictions allow us to achieve a robust and stable study of potential predictability of ENSO and investigate the decadal/interdecadal variation of ENSO potential predictability.

In this paper, we will explore the information-based and ensemble-based potential predictability for ENSO using the long-term retrospective ensemble predictions from 1856-2003 with the LDEO5 model. Relationships between actual prediction skill measures and potential predictability measures at various time scales from interannual to decades will be investigated. Some theoretical properties of the information-based predictability measures will be examined by a long-term LDEO5 ensemble prediction over 100 years. With such a long-term retrospective ENSO ensemble prediction, some new findings and understandings in ENSO

predictability can be expected.

This chapter is structured as follows: Section 5.2 briefly describes the LDEO5 model and the method of ensemble construction. Section 5.3 gives the definitions of the actual skill measures, ensemble-based potential predictability measures, and information-based potential predictability measures. Section 5.4 examines the information-based potential predictability measures using the 148-yr retrospective ensemble prediction, in comparisons with previous findings found in Tang et al. (2008a). Section 5.5 discusses the relationship of information-based potential predictability measures and actual prediction skill at different time scales. Section 5.6 discusses the control factor of information-based potential predictability measures and followed by the summary in section 5.7.

5.2. The Strategy of Ensemble Construction

The strategy of ensemble construction in this study attempts to reflect two major sources of uncertainties in ENSO prediction: uncertainty in the initial condition and external stochastic atmospheric noise during the forecast period (e.g., Moore and Kleeman 1998). Thus, a joint perturbation, composed of the leading singular vector of SST (SV1_sst) perturbation at initial condition and the leading stochastic optimal perturbation of winds (SO1_wind) during the whole forecast period (denoted by SV1_sst +SO1_wind), was applied to construct ensemble predictions. The SV1_sst and SO1_wind represent the optimal growth of perturbation due to uncertainties in both initial SSTA and atmospheric transients, respectively. They were obtained

by perturbing the tangent linear model (TLM) of the LDEO5 model outlined in chapter 2 and 3. It was found that this joint perturbation is able to provide reliable and high resolution ENSO probabilistic forecasts by the LDEO5 model, where the reliability was validated by the reliability diagram (i.e., measuring the consistency between forecast distribution and the corresponding observed distribution) and the resolution (i.e., the difference between observed relative frequency and probability of the climatological forecast) was measured by the Brier skill score (BSS; e.g., Wilks 1995), RPS (the ranked probability score; Epstein 1969; Murphy 1969; Murphy 1971), and RPSS (the ranked probability skill score; Wilks 1995). The details of the ensemble construction by this joint perturbation and resultant probability prediction can be found in chapter 4.

The model is initialized by the assimilation of SST every month for 1856-2003 (Chen et al. 2004), thus a total of $148 \text{ years} \times 12 \text{ months/year} (=1776)$ forecast initial conditions were obtained. From each initial time, an ensemble forecast was performed with the ensemble size (M) of 100, and leading time of 24 months. Thus, there are a total of $1776 \times 100 \times 24$ (4262400) forecasts for the ensemble experiment.

5.3. Prediction skill Metrics

5.3.1 Metrics of Actual Prediction Skill

The correlation-based skill and MSE-based skill are used to measure deterministic prediction skill. The overall skill of ensemble mean predictions over the 148 yrs is measured by anomaly correlation (R) and root mean square error (RMSE) (e.g., Scherrer et al. 2004). For an individual prediction, its skill is evaluated by the correlation and MSE, calculated using the individual forecast of the entire lead time of 24 months, called correlation and MSE of individual prediction (CIP and MSEIP), as in Moore and Kleeman (1998) and Tang et al. (2008a). In this study, the predicted Nino3.4 SSTA index (averaged over 5°N to 5°S, from 170°W to 120°W) and its observation counterpart are used to evaluate the ENSO prediction skill.

5.3.2 Ensemble-based Measures of Potential Predictability

Ensemble mean (EM), ensemble spread (ES), and ensemble ratio (ER) are common ensemble-based measures of potential predictability that do not make use of observations..

$$EM(i, t) = \frac{1}{M} \sum_{m=1}^{M=100} T^P(i, t, m) \quad (5.1)$$

$$ES(i, t) = \sqrt{\frac{1}{M-1} \sum_{m=1}^M [T^P(i, t, m) - EM(i, t)]^2} \quad (5.2)$$

$$ER(i, t) = \left| \frac{ES(i, t)}{EM(i, t)} \right| \quad (5.3)$$

where EM , ES , and ER are a function of initial time i and lead time t . M is ensemble size, i.e. 100 here. T is the index of Niño3.4 SSTA, the superscripts p and o denote predictions (forecasts) and observations respectively; N is the number of initial conditions used ($N=1776$). Note that instead of using the EM, the square of ensemble mean, denoted by EM2, is used as a potential skill measure in this study since it is a better indicative of the magnitude of ENSO signal as suggested in Tang et al. (2008a).

The information-based predictability measures are introduced in Chapter 1.3. The actual prediction measures and potential predictability measures used in this study are summarized in Table 5.1, as functions of initial time, lead time, or both. R, RMSE, CIP, MSEIP are actual prediction skill measures because they depend on observation whereas EM2, ES, ER, RE, PI, and PP are potential predictability skill measures that do not involve observation data.

5.4. Characteristics of Information-based Measures of the LDEO5 Model

5.4.1 The general characteristics of RE, PI, and PP

We first explore the properties of the information-based potential predictability measures through the long-term ENSO ensemble forecast with the LDEO5 model. Fig. 5.1 shows the information-based measures RE, PI, and PP at the Nino3.4 SSTA index, as a function of lead

time and initial condition, for the time period from 1960 to 2003¹. In Fig. 5.1, several features are displayed similar to those presented in Tang et al. (2008a), where a shorter period ENSO hindcasts from 1981-1998 were obtained by two hybrid ENSO models. These common features are (i) Large RE peaks are related to strong ENSOs. For example, strong El Niños that occurred in 1972/73, 1982/83, and 1997/98, and La Niñas in 1974/75, 1988/89, and 1999/2000 can be identified by the strong peaks in the RE plot (Fig. 5.1a). On the other hand, it is difficult to connect ENSO variability with PI or PP since large PI and PP occurred frequently in Figs. 5.1b-c. (ii) RE declines significantly as lead time of prediction increases, whereas PI and PP display relatively smooth variations with lead time and initial conditions. Mathematically, according to definitions of RE, PI, PP in (1.15)-(1.17), a primary explanation for the reduction of these information-based predictability with lead time is related to decreasing of the signal component (SC) or decreasing the dispersion component (DC, which is inversely determined by the ES). These features in Fig. 5.1 are consistent with that presented in Tang et al. (2008a), confirming that the RE is a better indicator of ENSO variability than PI/PP. This is mainly due to the properties of these measures, i.e., RE depends more on the signal component SC while PI or PP depends only on dispersion component DC. We will give detail discussions on this point in section 6. In addition, it has been argued that if the initial signal of the prediction is strong, more extra information can be provided from the forecast comparing with the climatological forecast, leading to a more skillful and reliable forecast (Kleeman 2002; Tang 2005; Tang

¹ We performed the analysis for the entire period from 1856-2003 but only plotted the short period from 1961-2003 in Fig. 1 just for a clearer presentation.

2008a).

An interesting feature of the PP/PI can be found when the lead time of prediction is increased to 24 months in Figs. 5.1e-f (the two bottom plots in Fig. 5.1), which was absent in Tang et al (2008a) where they only explored results within 12-month lead times. As seen in Figs. 5.1e-f, the large and small PI/PP slopes occur annually, indicating a strong seasonal variability of the PI and PP. Along each slope, the PIs/PPs started from different initial time (i) and varied with different lead time (t). However, PIs/PPs are actually corresponded to the same verification time [the time at the end of the forecast or the target time (v), i.e. $v=i+t$]. The occurrences of the slopes in the PI/PP plot are related to the ensemble spread ES. As defined in Eq. (1.15) and (1.17), PI or PP depends only on the standard deviation of forecast (σ_p) or the ensemble spread ES. This point can be verified by Fig. 5.2a, where similar feature of seasonal variability indeed exists in the ES plot. Large ESs occur at target times between boreal autumn and winter, while small ESs occur at spring and summer. This strong seasonal variability characteristic of ES can be more clearly indicated in Fig. 5.2b, where the x-axis is target month. A similar feature was presented in Karspeck et al. (2006) when they studied variations of the ES by an older version of the ZC model (version LDEO4). They suggested that the seasonal variability of ES is consistent with the seasonal cycle in ENSO signals. More precisely, a large ES occurs in the boreal autumn and winter when the SST variability is strong in the equatorial Pacific Ocean, whereas a small ES occurs during the boreal spring and early summer with a weaker SST variability. They suggested that ES can be better presented as a function of the calendar month

at the target month of the forecast (regardless of the lead time) than as a function of the initialization month.

An inverse relationship between ES and ENSO initial signals has been reported in some previous works, which was explained by the delayed oscillator mechanism (e.g., Cai et al. 2003; Zhou et al. 2008), i.e. a large (small) error growth occurs with weak (strong) signals at the initial time. To further see this, we binned ES according to the value of Nino3.4 SSTA index at initial times in Fig. 5.3a. As can be seen an inverse ES – signal relationship can be observed for lead times until 18 months, which is consistent with previous studies. However, if ES is grouped by SSTA at the target time, opposite features appear as displayed in Fig. 5.3b, i.e., ES increases with the magnitude of the Nino3.4 SSTA. The strong target-time-dependent features are also found in actual prediction skills such as anomaly correlation and RMSE as shown in Figs. 5.4a-b, where small correlation and small RMSE skills occur during the target times in the boreal spring and summer while both large correlation and RMSE skills appear during the fall and winter. A target-seasonal-dependent feature of correlation skill was also reported by Saha et al. (2006) using the National Centers for Environmental Prediction (NCEP) climate forecast system (CFS), where the SST forecasts show lower correlation skills from April to August with a significant correlation skill dropping at the target month in April, corresponding to the well-known “spring barrier” phenomenon. In the LDEO5 model, although the spring barrier was found relatively weak in the 148-yr retrospective forecasts (Chen et al. 2004), it still exists in our ensemble mean forecasts. As shown in Fig. 5.4a, higher correlation skills occur during

boreal autumn and winter, and the correlation skill drops sharply in April or May, and especially true for the lead times less than 12 months.

In Fig. 5.4b, however, the spring barrier is not shown in the RMSE skill. Instead, small RMSEs occur during the target time of the summer. This is most probably related to the feature of the seasonal variation of SSTA variability and its association with the magnitude of RMSE. Usually the Nino3.4 SSTA has relatively small variability (variance) in spring and summer, leading to relatively small RMSEs intrinsically. To remove this inherent impact from climatological variance, we use the “relative RMSE” measure (i.e. the ratio between RMSE and the RMSE of the climatological prediction) instead of RMSE, and obtained Fig. 5.5. Clearly, Fig. 5.5 shows significant increasing of the relative RMSE during the boreal spring, which is very consistent with the correlation skill in Fig. 5.4a.

Similar target-time-dependent features are also presented more or less in the ensemble-based and information-based predictability measures in Figs. 5.4c-f. The high similarity between PP and ES is as expected because that PP depends on ES and they have an inverse relationship according to the definition of PP. Thus, the PP/ES has the strong target dependent feature as RMSE. A smaller/larger EM2 is related to the ENSO signals at the target season, which may explain the smaller/larger correlation skill in Fig. 5.4a, because the anomaly correlation skill is often associated with ENSO signals. Among these potential skill measures, the RE shows unique features from others. RE is not very sensitive to the target season as the others, and it does not have very strong spring barrier. These features can be explained by the

definition of RE, where the common target-season-dependent features of EM2 and ES in Fig. 5.4c and Fig. 5.4d could be efficiently removed by the “ratio” operator.

It is a widely-used strategy to study ENSO predictability as a function of initial time as presented in literatures. The above results suggest the importance of the target time of prediction in ENSO predictability. In particular, the target-time-dependent feature displayed here might not be a model-dependent as reported in CFS ensemble prediction (Saha et al. 2006). A further analysis using more ENSO models is required for this issue.

5.4.2 Characteristic of Mutual Information

We return to the discussion of the overall features of information predictability measures. According to information theory, the averaged value of RE and PI over a long time period should be close to each other, approaching to an overall potential predictability measure, called the mutual information (MI). MI, as a function of lead time, is an indicator of the overall predictability of the target variable in a forecast model. This property is held well in present ensemble predictions of the LDEO5 model. In Fig. 5.6a, two sets of MIs that were computed from the averaged RE and PI respectively indeed have very close values at any given lead times. In addition, the MIs decrease smoothly as lead time increases, holding the monotonical property perfectly, and asymptotically approach to the minimum value. At lead times of 9-month and longer, the MIs varies very smoothly, suggesting ENSO predictability quickly approaches the minimum at lead times around 9-12 months and it keeps at that level until 2-yr for the LDEO5

model.

Figs. 5.6c-d show strong relationships between MI and the actual prediction skill (anomaly correlation skill and RMSE skill). Large MIs are related to small RMSEs but high correlation skill, and vice versa. The monotonical property of MI and its good relations with actual prediction skills indicate that MI is a good indicator of overall predictability.

For Gaussian variables, MI follows a theoretical relationship with correlation skill R as below (Desole 2004),

$$MI_{th} = -0.5 \ln(1 - R^2). \quad (5.4)$$

To examine (5.4) for the LDEO5 model, the MI obtained from the averaged RE and PI are compared with the estimated MI from actual R . As seen in Fig. 5.6b, except for the three largest MIs (corresponding the lead times of 1-3 months), the estimated MIs are very close to the MIs from the averaged RE/PI. The scattered points are distributed along the “perfect” diagonal line from lead times of 4-24 months. The reason of difference between potential MI and estimated MI at short lead times is not clear, but the largest three MI points are related to the high correlation skills of short lead times (1-3 months), at which the ensemble spread may be not fully developed, leading to a large RE/PI. A comparison of the LDEO5 model with the hybrid models in Tang et al. (2008a) reveals that the LDEO5 model provides a much better validation for (5.4).

Eq. (5.4) can be used to estimate a correlation skill (R_{th}) when observations are not available. Fig. 5.7a is the scatter plot between the actual correlation R and R_{th} obtained using the averaged RE/PI. Compared with the perfect R_{th} - R relationship (i.e. the diagonal line in Fig. 5.7a), the averaged RE gives a better R_{th} - R relationship than the averaged PI, especially at lead times longer than 15 months when there is low predictability. However, the estimated correlation skill R_{th} by the averaged RE always offers a larger skill than the actual correlation R , and meanwhile, the estimated correlation skill R_{th} by the averaged PI tends to underestimate the actual predictability, both estimated relationships departing from the perfect relationship (the diagonal line) but almost distributing symmetrically around the diagonal line at lead times less than 15 months. Thus, a revised R_{th} - R relationship is suggested here by averaging the two R_{th} s from RE and PI. The new R_{th} - R relationship is shown in Fig. 5.7b. As can be seen, the revised R_{th} - R relationship is almost perfectly along the diagonal line at the lead time 4-18 months, providing a good potential correlation skill for the LDEO5 model. For short lead times of 1-3 months, the R_{th} - R relationship is not good, probably due to relatively small ensemble spread at the beginning of prediction.

5.5. Relationship of Potential Predictability Measure and Actual Prediction Skill

5.5.1 The Relationship on the Decadal/Interdecadal Time Scale

In the proceeding section, we examined some important features and properties of information-based measures as well as their capability in characterizing actual prediction skill. In this section, we further examine the relationship of information-based measure (RE, PI, and PP) and actual prediction skill (CIP and MSEIP) on the time scales from interannual to decades. The identified relationship may have a theoretical contribution to ENSO predictability study, and a practical significance in estimating the confidence that we can place in future predictions using the same ENSO forecast model.

A running window of 25-yr was applied to the information-based measures and the actual prediction skill measures, namely that, they were evaluated at each window of 25 years, starting from January 1856 and moving forward by 1 month each time until December 2003. As can be seen in the left panel of Fig. 5.8, on the decadal/interdecadal time scale, there is an in-phase relationship between the RE/PI/PP and the correlation skill (CIP) with a correlation coefficient of 0.88/0.36/0.45, respectively. The best in-phase relationship is between RE and CIP, suggesting RE to be a good indicative of actual correlation skill at decadal/interdecadal scales. The strong relationship between RE and CIP is due mainly to the large contribution of signal component to RE, which is absent in PI/PP. The significant role of ENSO signal in prediction

skill has been addressed in Tang et al. (2008a), in which they found that correlation skill is associated with the variations of ENSO signals for the 100 year from 1881-2000. For example, the strong ENSO signal time periods such as the ends of 19th century and the late two decades of the 20th century have higher correlation skill, whereas the weak ENSO signal periods have low correlation skill.

On the other hand, the MSEIP has weak relations with the information-based predictability measures at the long time scales, especially with PI/PP, as seen in the right panel of Fig. 5.8. A poor relationship of PI/PP and MSEIP is seen which is mainly due to their strong time-scale-dependent nature; we will discuss this point in the next section (i.e. they have strong relationships only at the shorter interannual time scales in Fig. 5.10 and Fig. 5.11d).

The running mean method used above may not be able to present objectively a full spectrum of the relationship between RE/PI/PP and predictability; for example, the relationship is probably sensitive to the length of running window. To further explore the decadal/interdecadal relationships, we extract low-frequency components (>10 years) using the fast Fourier transform (FFT) filter. Shown in the left panel of Fig. 5.9 are the decadal/interdecadal variations of the low-frequency components of RE/PI/PP along with the correlation skill CIP. In-phase relationships can be seen between RE/PI/PP and CIP, with positive correlation coefficient of 0.68/0.4/0.4, respectively. Similar to Fig. 5.8a1, when the REs were large during the late 19th and the late 20th centuries, the correlation skill CIP were high; while when REs were small during time periods from 1910-1955, the correlation skill scores

were low. Compared with PI and PP, the RE has the closest in-phase relationship with the correlation skill CIP. On the other hand, the right panel of Fig. 5.9 depicts the weak/poor relationship of RE/PI/PP and MSEIP on decadal/interdecadal time scale. In addition, FFT gives an in-phase RE-MSEIP with correlation coefficient of 0.21, which is opposite to the Fig. 5.8a2 using the running mean method. It further implies the uncertain relation between RE and MSEIP on the longer time scale. Therefore, Fig. 5.8 and Fig. 5.9 indicate that the RE has a good in-phase relationship with the correlation skill CIP on the decadal/interdecadal time scales, and all information-based potential measures do not have clear relation with the MSEIP skill measure on the long time scales.

5.5.2 The Relationship on Interannual Time Scales

A further analysis explores whether such a good in-phase relationship of RE and CIP exists at interannual time scales and for individual forecast cases. Shown in Fig. 5.10 is the scatter plot of RE/PI/PP against CIP and MSEIP, where a 2-7 yr FFT filter has been applied to all variables to extract their interannual variability. Fig. 5.10 shows significant in-phase relationships between RE/PI/PP and CIP with correlation coefficient of 0.46/0.34/0.35; and significant inverse relationships are seen between the RE/PI/PP and MSEIP with correlation coefficient -0.36/-0.36/-0.38, respectively. Compared with the poor relationship between MSEIP and RE/PI/PP in Figs. 5.8-9, Fig. 5.10 suggests that the MSE-related relationships are highly time-scale-dependent. On the interannual time scale, RE/PI/PP is a good indicator of prediction skill. A large RE/PI/PP is associated with a large correlation skill and a low MSE

skill, and vice versa.

5.5.3. The Relationship on All Time Scales

On all time scales from seasonal to decades, scatter plots of information-based measures (RE, PI, and PP) and actual deterministic measures (CIP and MSEIP) are given in Fig. 5.11 using all original samples without filtering. Again, among three information-based measures, RE has the best relationship with CIP with a significantly high correlation coefficient of 0.51. A comparison of these results between the LDEO5 model and two hybrid models in Tang et al. (2008a) reveals that the LDEO5 model offers a higher correlation coefficient between CIP and RE. In addition, there are still some uncertain relationships in Fig. 5.11a1. This uncertain relation is consistent with the “triangular relationship” found in previous studies (e.g. Tang et al. 2005, 2008a), namely, when RE is large, correlation prediction skill was high; whereas when the RE was smaller, the RE-CIP relationship has more uncertainties. A small RE is related to weak ENSO signal and large model error, which has a smaller signal-to-noise ratio and thus a lower predictability.

On all time scales, Fig. 5.11a2 also shows a triangular relationship between MSEIP and RE, i.e. when RE is large, the RMSE is small; whereas when RE is small, the MSEIP is uncertain. This uncertain relationship of RE-MSEIP is probably due to the fact that the RE-MSEIP relationship is scale-dependent. For example, they have a weak in-phase relation with correlation coefficient of 0.21 on the decadal/interdecadal time scale in Fig. 5.9a2, but an

inverse relationship at the interannual time scale in Fig. 5.10a2 with a correlation coefficient of -0.36. On the other hand, it's interesting to see an inverse relationship between MSEIP and PI/PP in Fig. 5.11b2 or Fig. 5.11c2 with a negative correlation coefficient of -0.27/-0.30 (even better than RE-MSEIP relation). In the following section we will further discuss the scale-dependent relationship in more detail by the cross-wavelet analysis.

5.5.4 Cross-wavelet Analyses For Potential Measures and Actual Measures

The Cross-wavelet transform (XWT) method is used to examine relationships between two time series in time-frequency space (e.g., Grinsted et al. 2004). From the XWT analysis, the common power and relative phase can be revealed. The phase differences between two variables are depicted by the direction of a vector, with in-phase pointing right, anti-phase pointing left, and the first variable leading the second by 90° pointing straight down. In this study, a continuous XWT technique with the Morlet wavelet as the mother function was applied. Monte Carlo methods are used to assess the statistical significance against a red noise background. The standard software package of cross-wavelet transform is available online⁺ (Grinsted et al. 2004). Further details on XWT analysis can be found in Grinsted et al. (2004) and Torrence and Compo (1998).

To further illustrate characteristics of relationships between information-based predictability (RE, PI) and actual skill (CIP and MSEIP) at various time scales, cross-wavelet analyses were

⁺ <http://www.pol.ac.uk/home/research/waveletcoherence>

performed and given in Figs. 5.12a-d. The PI has XWT Plots very similar to the PP thus it is not shown here. The thick contour in Fig. 5.12 encloses regions of greater than 95% confidence, using a red-noise background spectrum. Generally, in-phase relationships (arrow pointing right) can be seen for information-based measures RE/PI and CIP in Fig. 5.12a and Fig. 5.12c, respectively. They are well consistent with the proceeding results that RE is superior to PI/PP as an indicator of prediction skill. At time scales longer than 10-yr, a strong in-phase relationship is found in the RE–CIP plot over 100-yr (Fig. 5.12a1). On the interannual time scale, in-phase relationship of RE/PI-CIP is also seen, but the relationship varies significantly from time to time and seems related to the variability of ENSO signals. There is even no identified relationship during the weak ENSO period from 1920-1950 on the interannual time scale, while strong correlations occurred in the two strong ENSO signal periods (1880-1920 and 1960-2000), indicating that the RE–CIP relationship depends highly on ENSO signals.

In contrast to the consistent in-phase relationship between CIP and RE/PI at all time scales, the relationship between RE and MSEIP exhibits scale-dependent features (Fig. 5.12b). At interannual time scale, an out-phase relationship is often seen from period to period, while at decadal/interdecadal time scale, strong in-phase relationship occurred in the periods of 1930-1950 and 1960-1980. These opposite relationships at the short and the long time scales lead to the poor/weak overall relationship between RE and MSEIP as shown in Fig. 5.11a2. The PI-MSEIP XWT plot also have strong scale-dependent features, showing without close relationship at longer time scales but a strong inverse relationship at interannual time scales

Thus, in contrast to the overall poor RE-MSEIP relationship, scale-dependent relationships between PI/PP and MSEIP lead to a good overall inverse relationship. These results are consistent with Figs. 5.11b2-c2, where PI/PP has a stronger inverse relationship with MSEIP than RE at all time scales. Therefore, PI/PP probably is a better potential measure than RE in assessing the MSE skill for individual forecasts.

5.6. The Control Factor of Potential Measures

From proceeding sections, we found that RE has better relationships with CIP than PI or PP. To explain this good relationship, RE is decomposed into the signal component (SC) and dispersion component (DC) according to (18). A 2-yr running mean method was applied on the data focusing on examining the variations on longer time scales ($>2\text{yr}$). Fig. 5.13 shows the temporal variations of RE, SC, and DC during the 148 years. Decadal/interdecadal variability can be seen in time series of RE and SC while DC shows very weak variability after applying the 2-yr running mean. During the higher SC periods (i.e. the end of 19th century and the end of the 20th century), the signal component term SC is very larger than DC; whereas during the weaker SC period (i.e. 1910-1940), DC and SC have comparable contributions to RE. Therefore, the variations of phase and amplitude of RE are mainly determined by the signal part SC or EM2 or ENSO signals; but DC is still important when ENSO signals are weak. The significant contribution of ENSO signal to ENSO predictability is in agreement with many recent ENSO predictability studies (e.g., Tang et al. 2008a), that is, stronger ENSOs have the higher prediction skill. The fundamental reason for decadal/interdecadal ENSO predictability

are still not very clear, but the strength of ENSO signal is a key factor (e.g., Tang et al. 2008a).

To further illustrate the important role of signal component SC, temporal variations of RE, PI, and PP are displayed in Fig. 5.14 along with the temporal variations of ENSO signal defined alternatively by absolute value of the observed NINO3.4 SSTA index at the target time at a given lead time of e.g. 6-month lead. As can be seen, all ENSO events and temporal variations of ENSO signals can be well identified by RE, with a significant high correlation coefficient of 0.74 over the 148 yrs. Again, RE shows the closest relationship with signal than PI and PP, as indicated by its definition. Similar results can be obtained at other lead times (not shown).

Because SC and DC are two important factors that determines predictability measures, it is reasonable to classify all those measures in Table 1 into two groups: (i) signal factor (EM2, ER, RE, CIP) and (ii) dispersion factor (ES, PI, PP, MSEIP). High correlation coefficients can be found between signal factors and correlation skill, and between dispersion factors and MSE skill, as shown in Table 5.2.

The signal component SC or EM2 plays a key role in determining RE-CIP relationship. Without the contribution of ES, the EM2-CIP relationship in Fig. 5.12e can still keep the main features of the RE-CIP in Fig. 5.12a on the interannual time scale. However, if a measure of potential predictability consider both the DC and SC components, as indicated by RE and ER, it is better than one that only considers either DC or SC. In other word, a good potential measure should include joint contributions of signal and spread in assessing the correlation-based skill,

e.g. ER and RE.

For a MSE-based skill measure, due to the offsetting scale-dependent relationships of ER/RE and MSEIP as shown in Fig. 5.12, ER/RE could not be superior to PI/PP, especially on all time scales and for individual predictions.

5.7. Discussion and Summary

One important task of ENSO predictability study is to seek good potential predictability measures, by which the uncertainty of individual prediction skill can be estimated without involving observations. In this study, the newly developed information-based potential measures (RE, PI, and PP) and the classic ensemble-based potential measures (EM2, ES, and ES) are explored based on their capability in quantifying the actual prediction skill (CIP and MSEIP) using long-term ensemble predictions by the LDEO5 model. Emphasis was placed on stable and robust relationship between potential predictability and actual prediction skill at various time scales. The relationship has practical significance and offers a practical means of estimating the confidence level of individual predictions. The decadal/interdecadal relationship between potential predictability and actual predictability has not been addressed in previous studies of ENSO predictability due to the lack of sufficient long retrospective predictions.

In this study, the ensemble was produced by using the optimal initial perturbation of SV1_SST and the perturbation of the stochastic optimal winds field during the forecast period. It was found that other ensemble construction methods such as using either the SV1_SSTA or

the realistic stochastic wind could not offer reliable ensemble predictions by the LDEO5 model (chapter 3). From the analysis of the 148-yr ensemble prediction, a good in-phase relationship is found between relative entropy RE and correlation skill CIP at multi-time-scale (i.e. from interannual to decadal/interdecadal time scales). The mutual information is a good measure for overall prediction skill. Different from Weather forecast, the signal component in ENSO predictions is much stronger than the dispersion component (noise), thus the predictability is dominated by ENSO signals.

RE is determined mainly by the signal component SC in ENSO predictions, but the dispersion component DC can play an important role in the predictability of weak ENSO periods during which the SC is comparable with DC. In terms of comparison with the ensemble-based predictability measures, RE and ER have comparable relations with the actual correlation skill CIP, probably because both consider ENSO signal and noise (i.e. ensemble spread ES).

Through the cross-wavelet analyses, we confirm that RE has a consistent strong in-phase relationship with the correlation skill CIP at time scales from interannual to decades. RE has an in-phase relationship with MSEIP on the long decadal/interdecadal time scales, but shows an inverse relationship with MSEIP on the interannual time scales. Thus, the RE-MSEIP relationship is highly time-scale-dependent. On the other hand, PI/PP does not show significant scale-dependent features with MSEIP. Therefore, at all time scales and for individual forecasts, either PI or PP can be a better predictability measure than RE for the MSE-based skill. In a

practical predictability evaluation, all information measures RE and PI/PP should be explored because they characterizes actual prediction skill from different aspects.

One interesting result found in this study is that the prediction correlation skill, RMSE, ES and PP/PI at lead times less than 12 months highly depends on target time. We identified that similar features are consistent with those shown in the NCEP CFS model or the LDEO4 model. Usually the correlation, EM2, RMSE, and ES are relatively high when the target time is in boreal winter and fall, whereas the low skills occur at the target time of boreal spring and summer.

There exists the 'spring barrier' in LEDO5 prediction, i.e., the correlation skill drops significantly while prediction in the boreal spring. The spring barrier is not obvious in RMSE skill, simply because the ENSO variability (e.g., SSTA variance) often is weak in boreal spring, leading to small value of RMSE at this season intrinsically. This can be confirmed by using the relative RMSE, which removes the inherent impact of ENSO variability, to examine the MSE-based prediction skill as a function of target time. As shown in Fig. 5.5, the relative RMSE clearly has spring barrier feature like correlation skill. The spring barrier is also striking in potential predictability measures PI/PP/ES, if we remove the inherent impact of ENSO variability like RMSE (not shown). In contrast to the strong spring barrier phenomena in the actual predictability correlation and the relative RMSE, and the potential predictability measures PI/PP, ES, EM2, a weaker spring barrier is found in the RE, showing the advantage of including both the signal and noise components.

Cautions should be borne in mind. The results and findings present in this study are based on the LEDO5 model and the chosen metrics of skill, thus they might be model and metrics dependent. For example, we explored ENSO predictability using CIP and MSEIP to measure prediction skill in this study. When the chosen metrics have been widely used in the field of predictability study, they might not be able completely characterize all properties of predictability. These concerns need to be addressed through more comprehensive analyses. Nevertheless, this work is the first exploration of ENSO information-based potential predictability over the past one and half century, providing insights on ENSO predictability, especially offering a practical means to estimate the confidence level for individual forecasts.

Table 5.1 Summarization of prediction skill measures used in this study. Prediction skill measures are as functions of either lead time (t) or initial time (i), or both.

Actual skill measure		$R(t)$	$RMSE(t)$	$CIP(i)$	$MSEIP(i)$
Potential skill measure	Ensemble_based measure	$EM2(t, i)$	$ES(t, i)$	$ER(t, i)$	/
	Information_based measure	$RE(i, t)$	$PI(i, t)$	$PP(i, t)$	$MI(t)$

Table 5.2 Correlation coefficients between actual prediction skill and potential prediction skill over the 148 years (from January 1856 to December 2003)

	EM2	ER	RE	ES	PI	PP
CIP	0.51	0.51	0.51	-0.28	0.26	0.28
MSEIP	-0.05	-0.07	-0.06	0.29	-0.26	-0.27

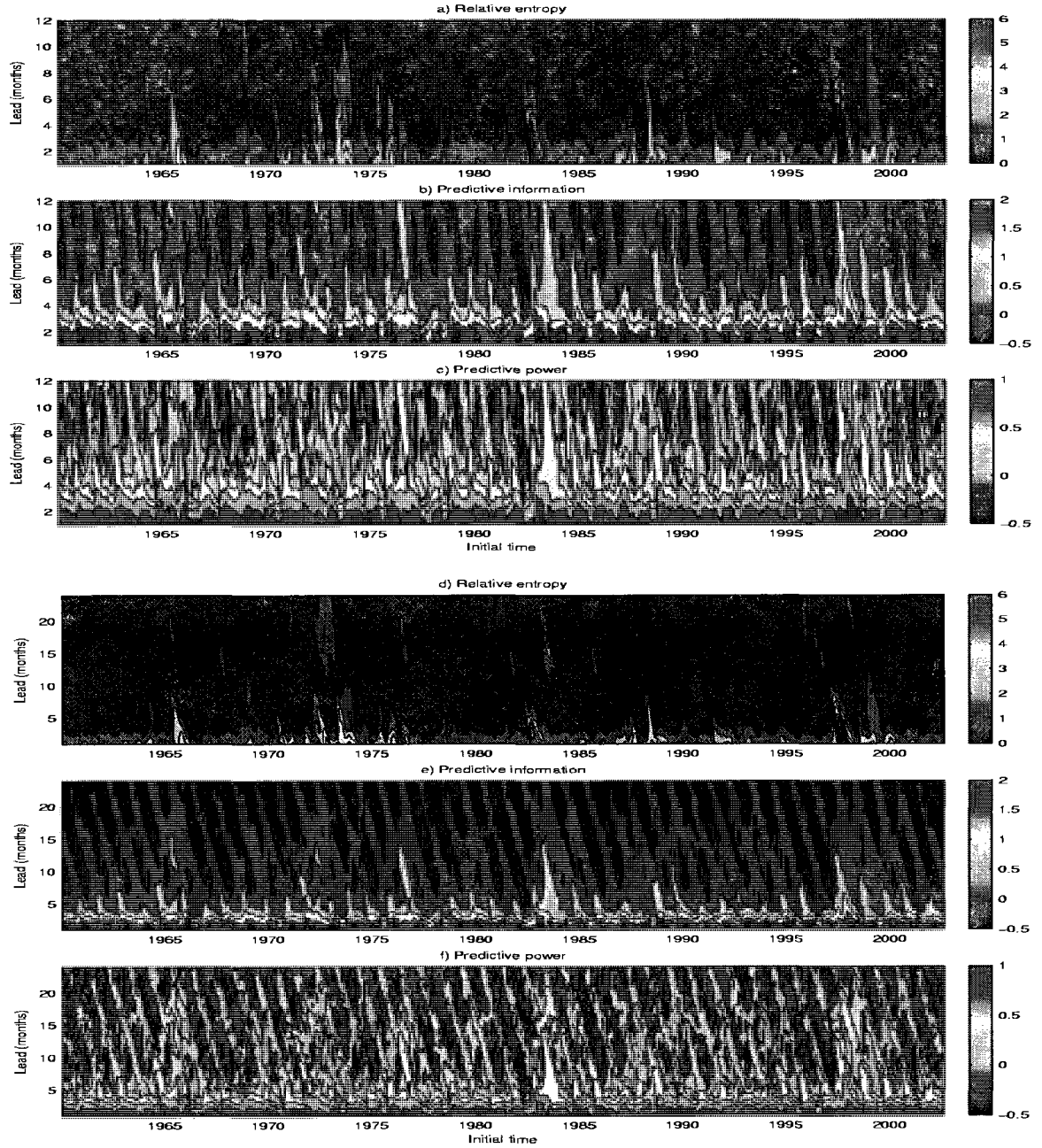


Fig.5.1. The left panel: a1) Relative entropy (RE), b1) predictive information (PI), and c1) predictive power (PP) of the Niño3.4 SSTA index as a function of initial time of each prediction and lead time (months), from Jan. 1960 to Dec. 2003 for the LDEO5 model. The right panel: same as the left panel, but for lead time from 1-24-month.

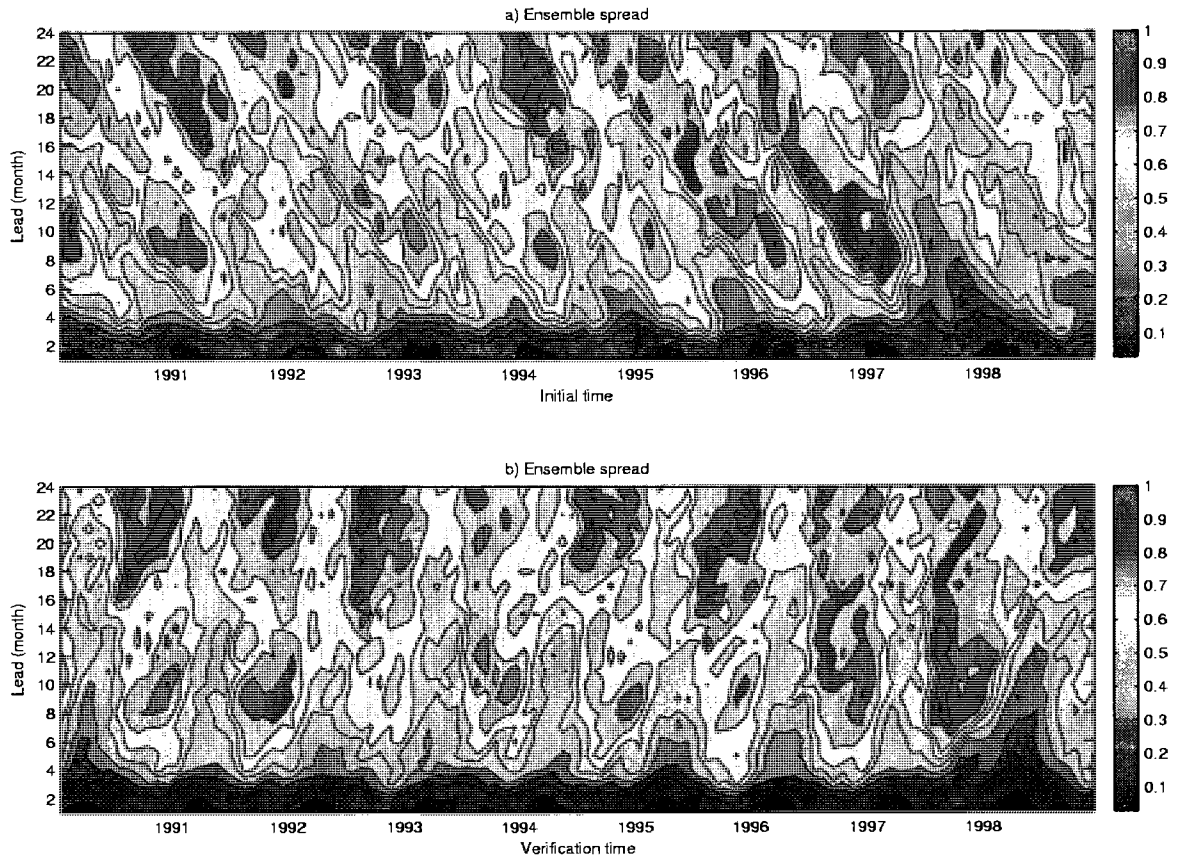


Fig. 5.2 a) Ensemble spread (ES) for time period from Jan. 1990 to Dec.1998 as a function of lead time (month) and initial time; b) same as a) but ensemble spread as a function of verification time.

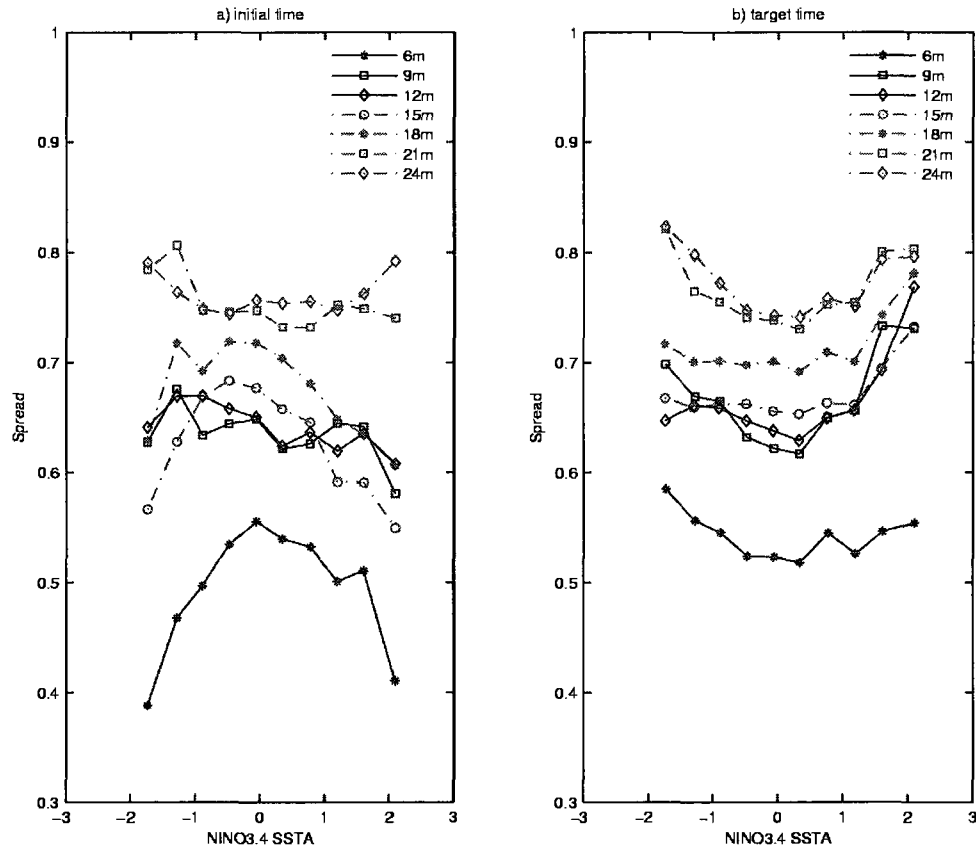


Fig. 5.3 a) Ensemble spread ($^{\circ}\text{C}$) as a function of NINO3.4 SSTA index at lead times of 6, 9..., 24 month, respectively. ES is binned by SSTA at the initial time. b) Same as a) but at the target time.

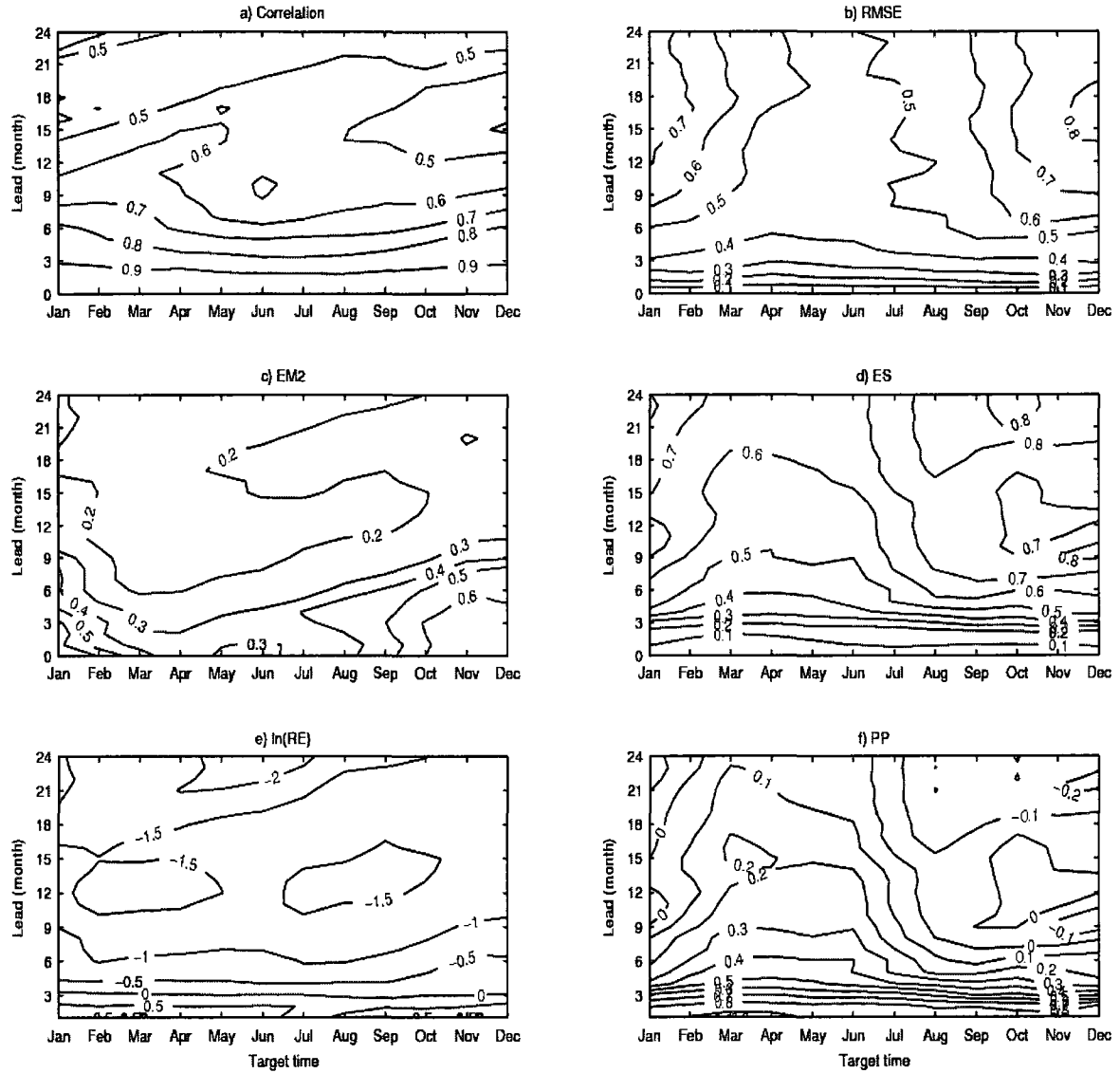


Fig. 5.4 a) Anomaly correlation of the LDEO5 ensemble mean forecasts of the monthly mean Nino3.4 SSTA over the 148-yr time period, as a function of target month (horizontal) and lead (vertical; in months); b) Same as a) but for RMSE; c) EM2; d) ES; e) $\ln(\text{RE})$; f) PP.

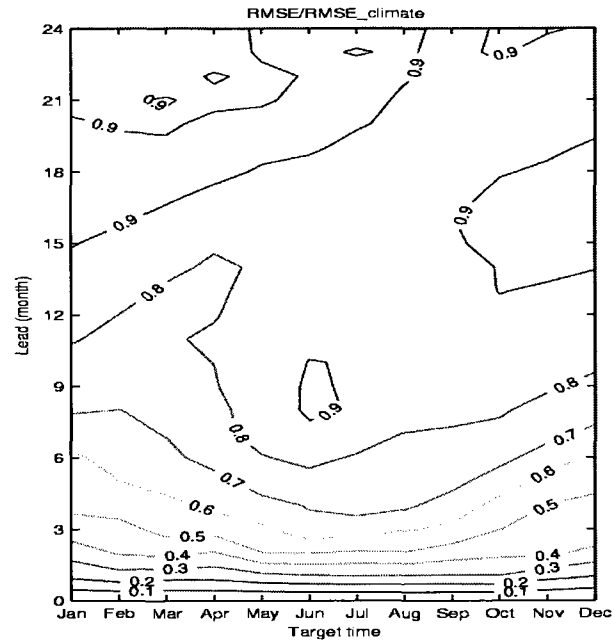


Fig. 5.5 Same as Fig. 5.4 but using a relative RMSE skill measure, defined by the RMSE over the RMSE of the climatological prediction. The climatological RMSE is as function of the calendar months.

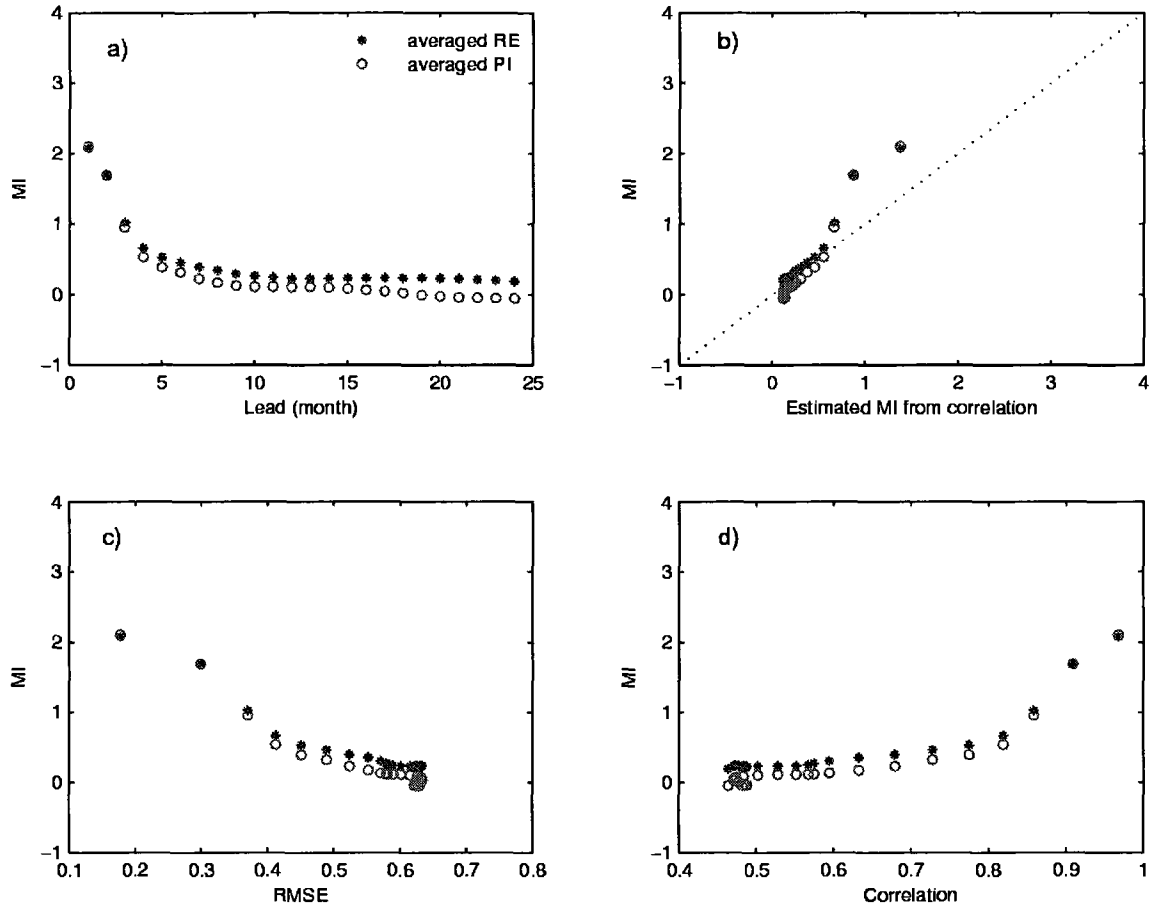


Fig. 5.6 a) Variation of MI calculated by averaged RE and PI over all initial conditions as functions of lead times; b) MI from averaged RE and PI against the estimated MI from correlation skill using Eq. (5.4). c) Averaged RE and PI versus RMSE skill; d) Averaged RE and PP versus correlation skill.

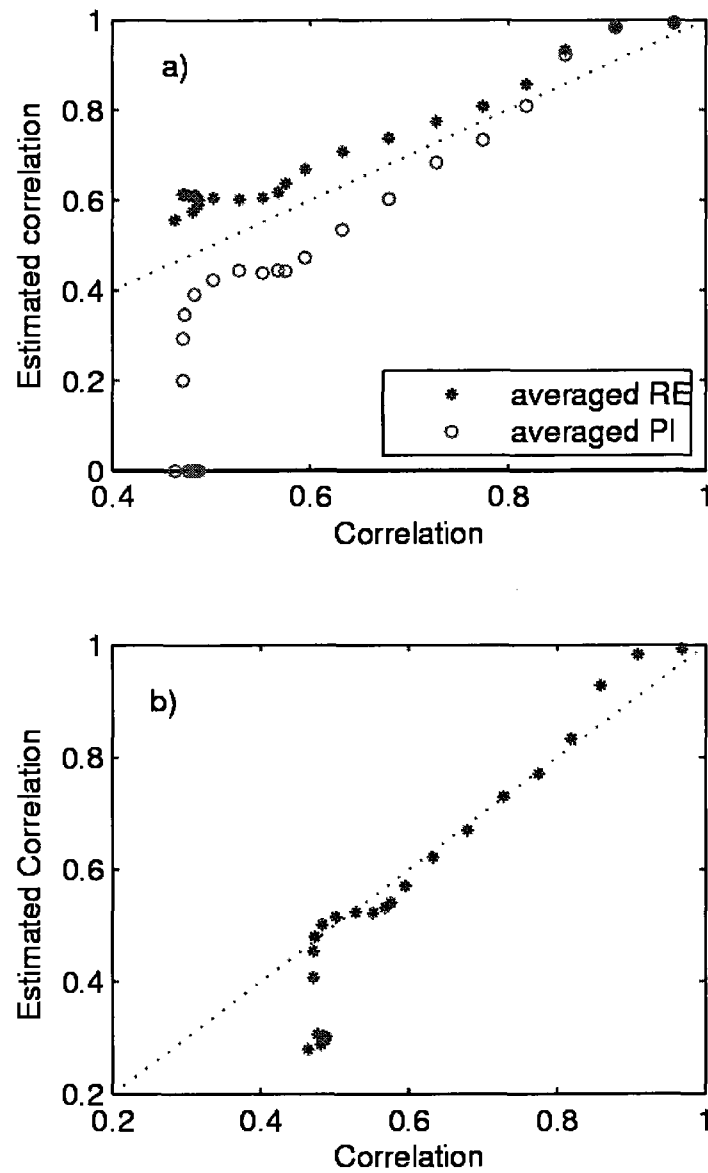


Fig. 5.7 The estimated correlation skill calculated by Eq. (5.4) against the actual skill. In a) MI is obtained by averaged RE and averaged PI, respectively and in b) MI is obtained by the mean values of the averaged RE and averaged PI.

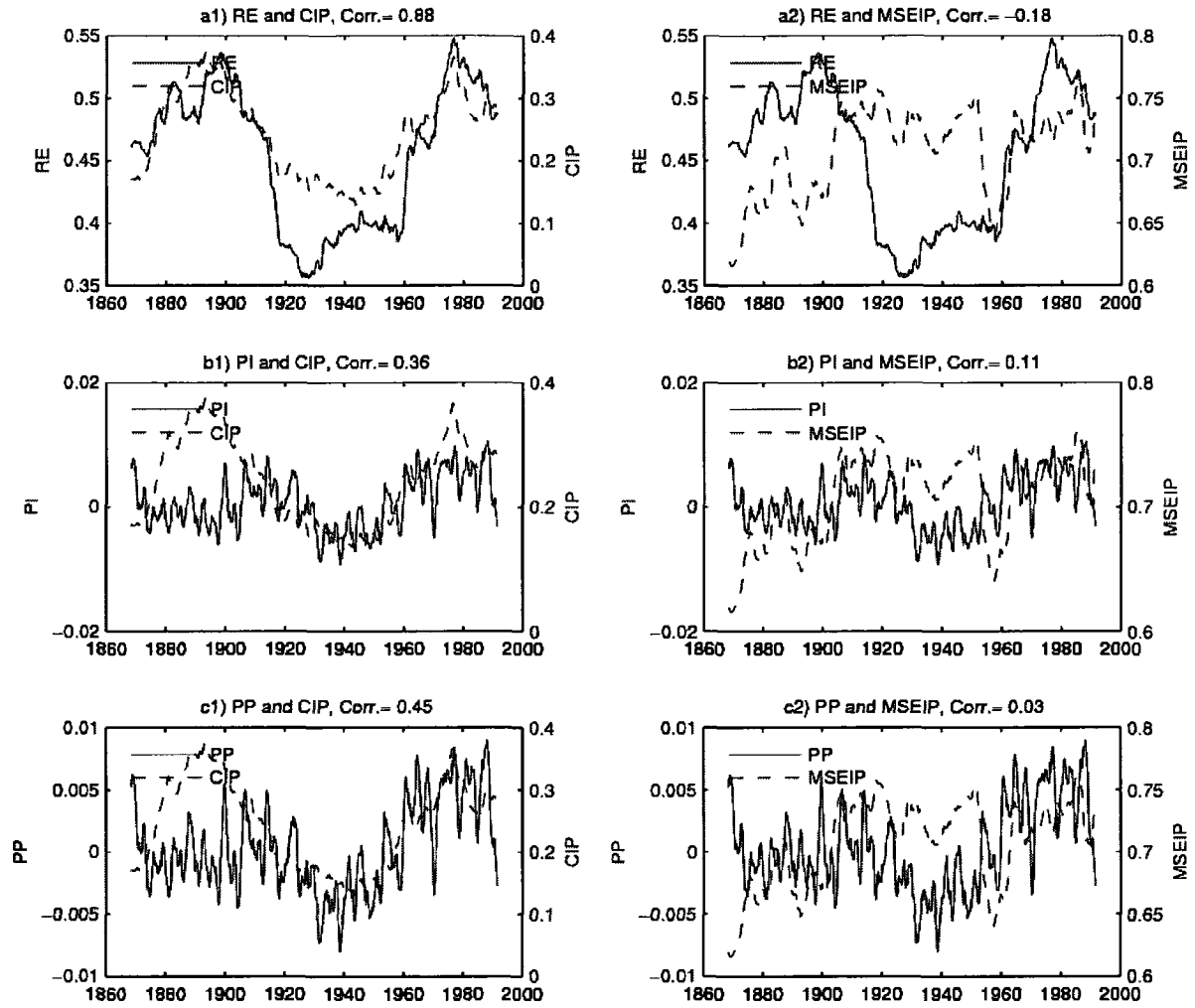


Fig. 5.8. Left panel: decadal/interdecadal variations of a1) RE and CIP; b1) PI and CIP; c1) PP and CIP. Right panel: decadal/interdecadal variations of a2) RE and MSEIP; b2) PI and MSEIP; c2) PP and MSEIP (A 25-yr running window has been applied for all data).

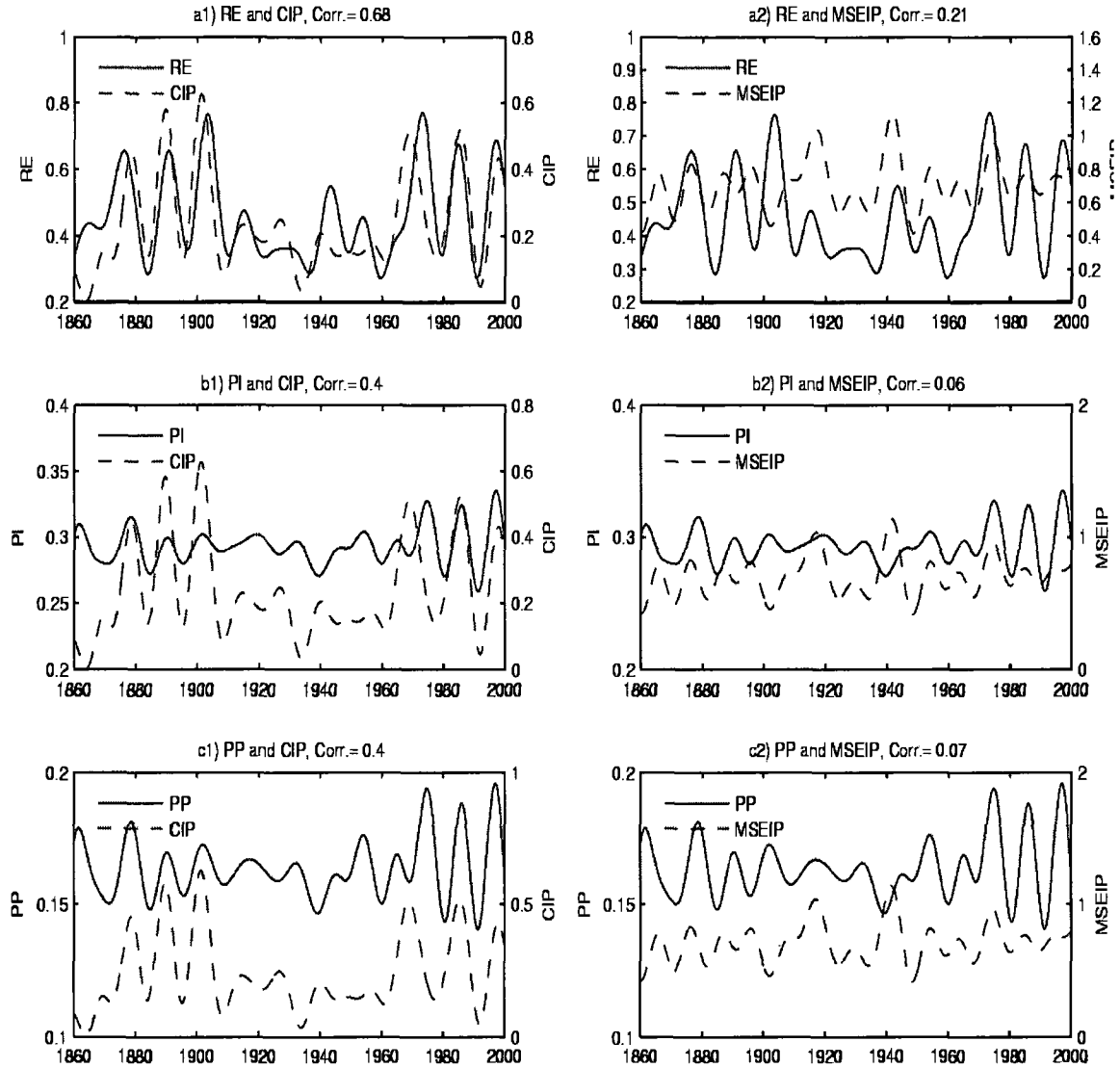


Fig. 5.9. Left panel: decadal/interdecadal variations of a1) RE and CIP; b1) PI and CIP; c1) PP and CIP. Right panel: decadal/interdecadal variations of a2) RE and MSEIP; b2) PI and MSEIP; c2) PP and MSEIP (A 10-yr low-pass FFT filter has been applied).

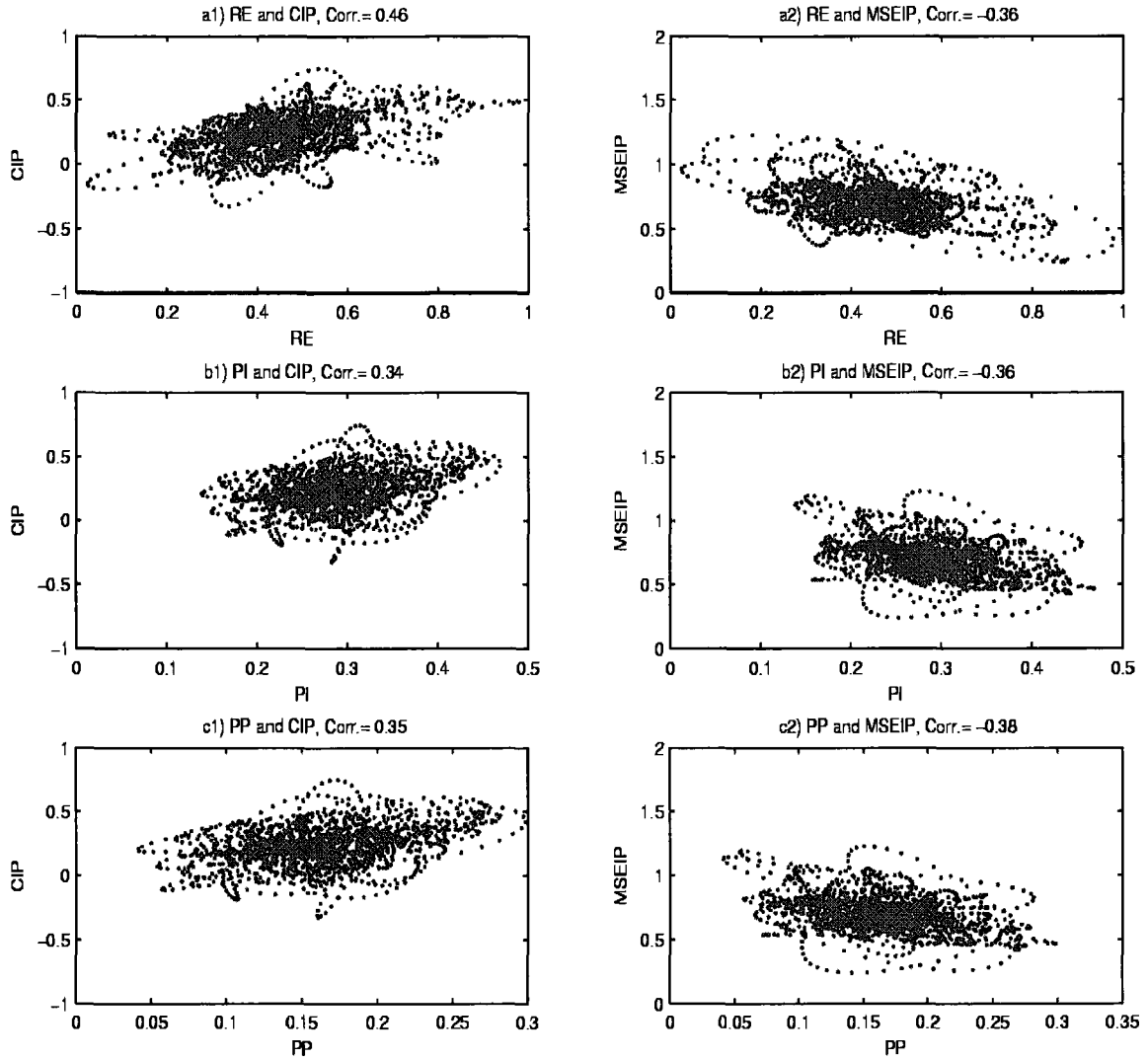


Fig. 5.10. Left panel: scatter plots of the individual correlation skill CIP against a1) RE; b1) PI; c1) PP. Right panel: scatter plots of the individual MSE-based skill MSEIP against a2) RE; b2) PI; c2) PP (at interannual time scales using a 2-7-yr FFT filter).

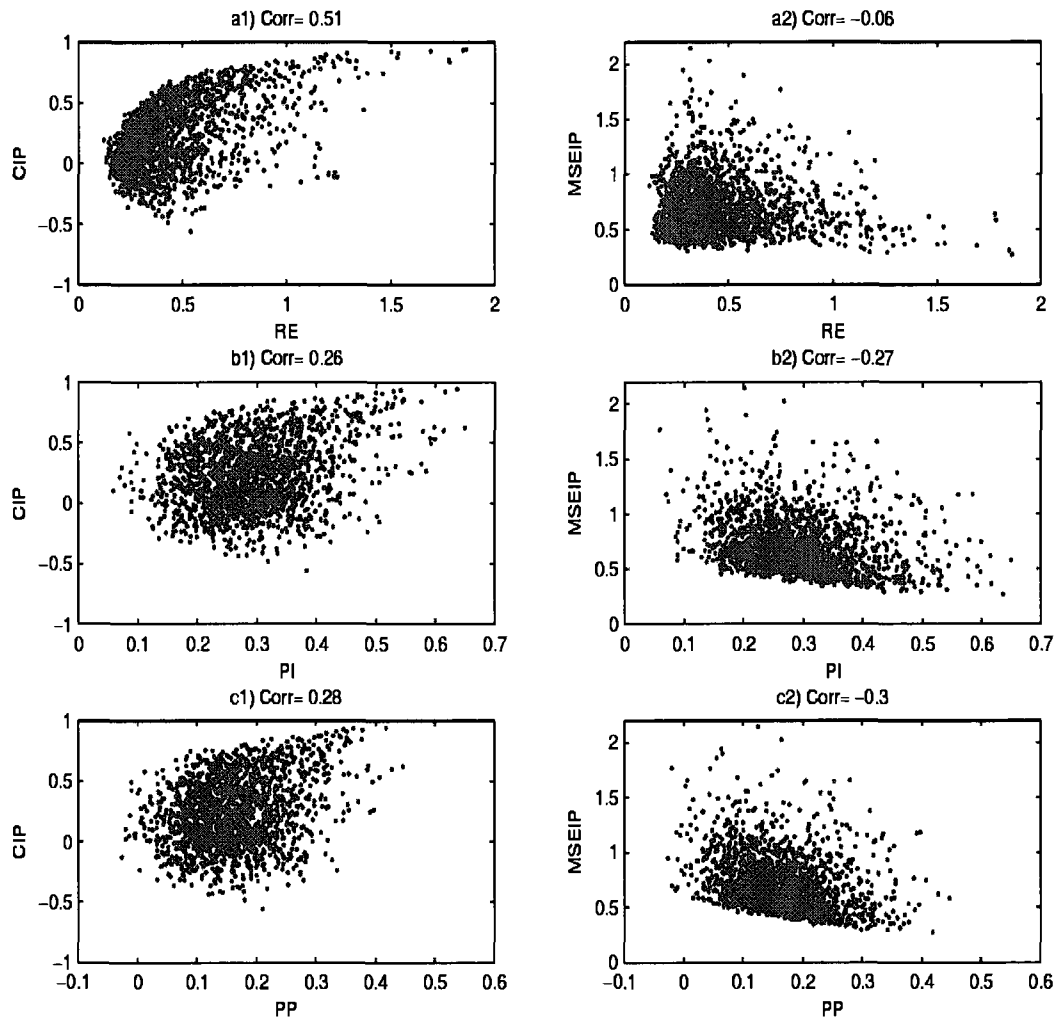


Fig. 5.11 Left panel: the correlation skill CIP against a1) RE; b1) PI; c1) PP. Right panel: MSEIP against a2) RE; b2) PI; c2) PP.

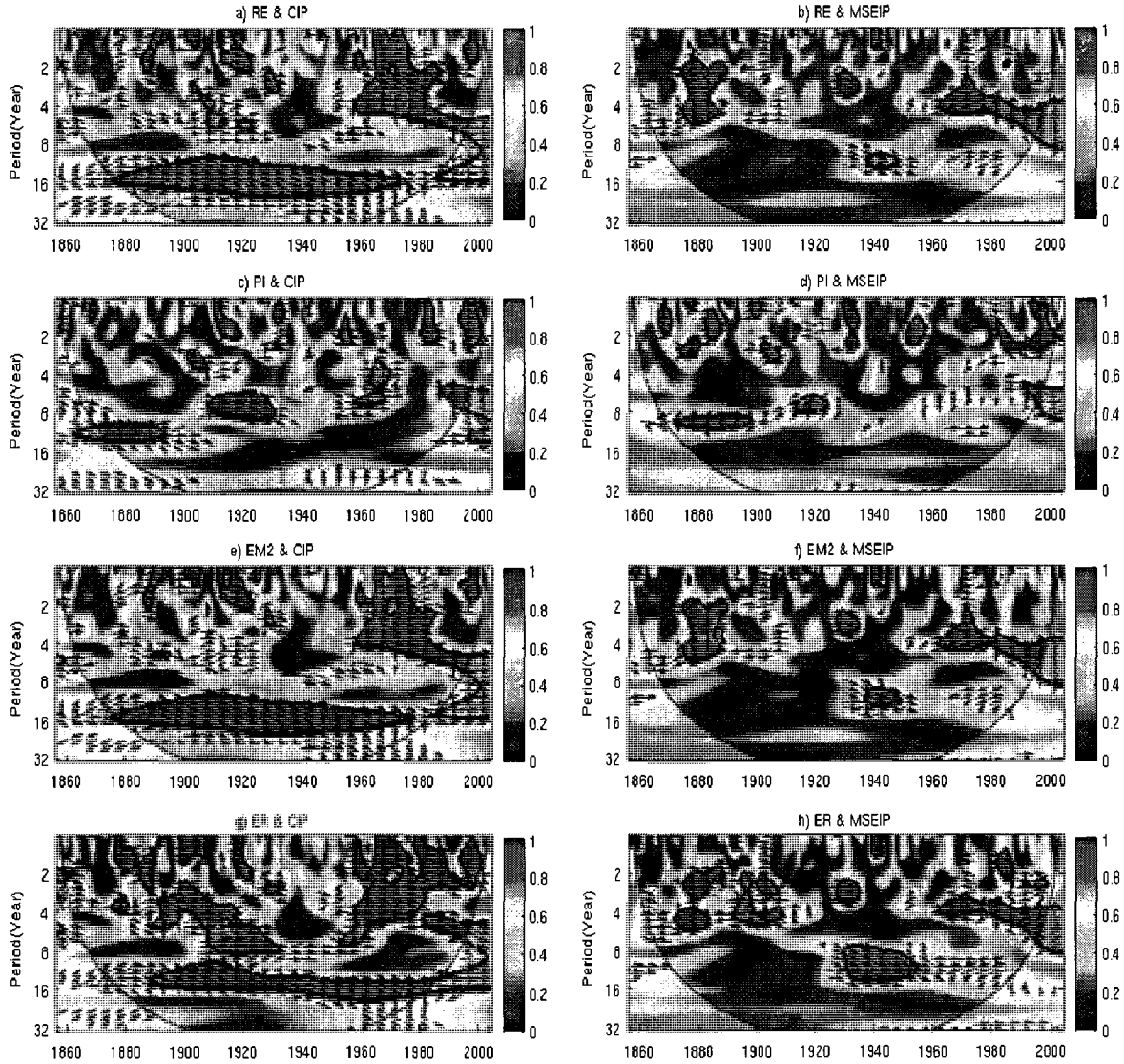


Fig. 5.12. The cross-wavelet analysis for i potential measures and deterministic prediction skill measures. a) RE and CIP; b) RE and MSEIP; c) PI and CIP; d) PI and MSEIP; e) EM2 and CIP; f) EM2 and CIP; g) ER and CIP; h) ER and MSEIP. The thick contour encloses regions of greater than 95% confidence, using a red-noise background spectrum. The relative phase relationship is shown as arrows, with in-phase pointing right, anti-phase pointing left and information-based measures leading actual skill measures by 90° pointing straight down.

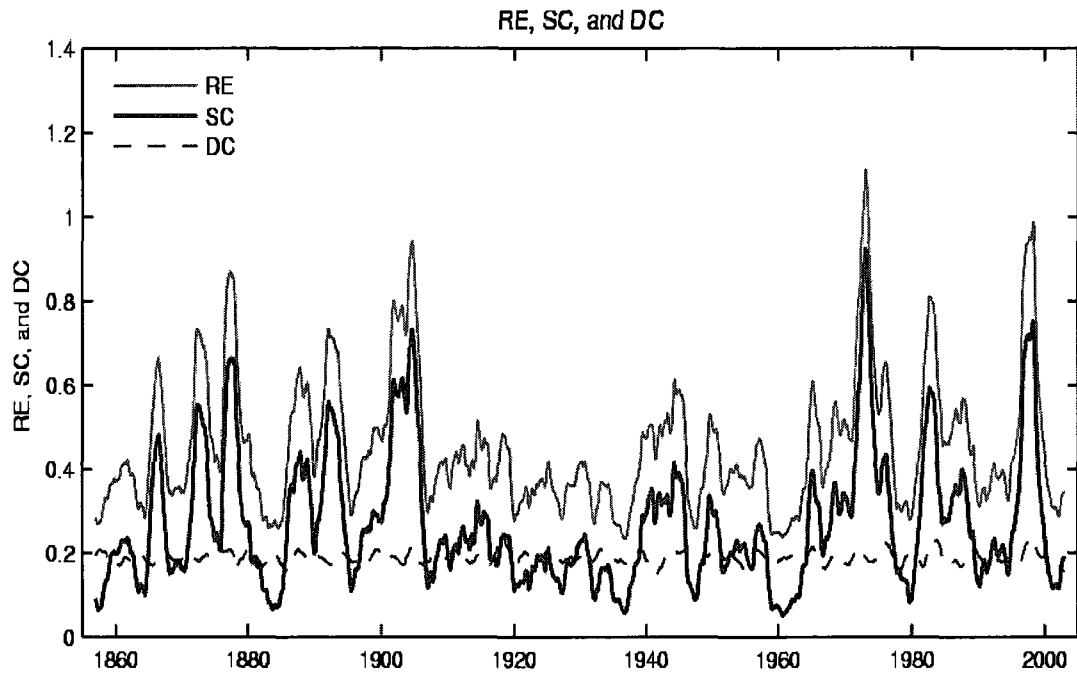


Fig. 5.13. Temporal variations of relative entropy (RE; solid), the signal component (SC; solid thick line) and the dispersion component (DC; dash) of RE. (A 2-yr running mean method was applied on the data).

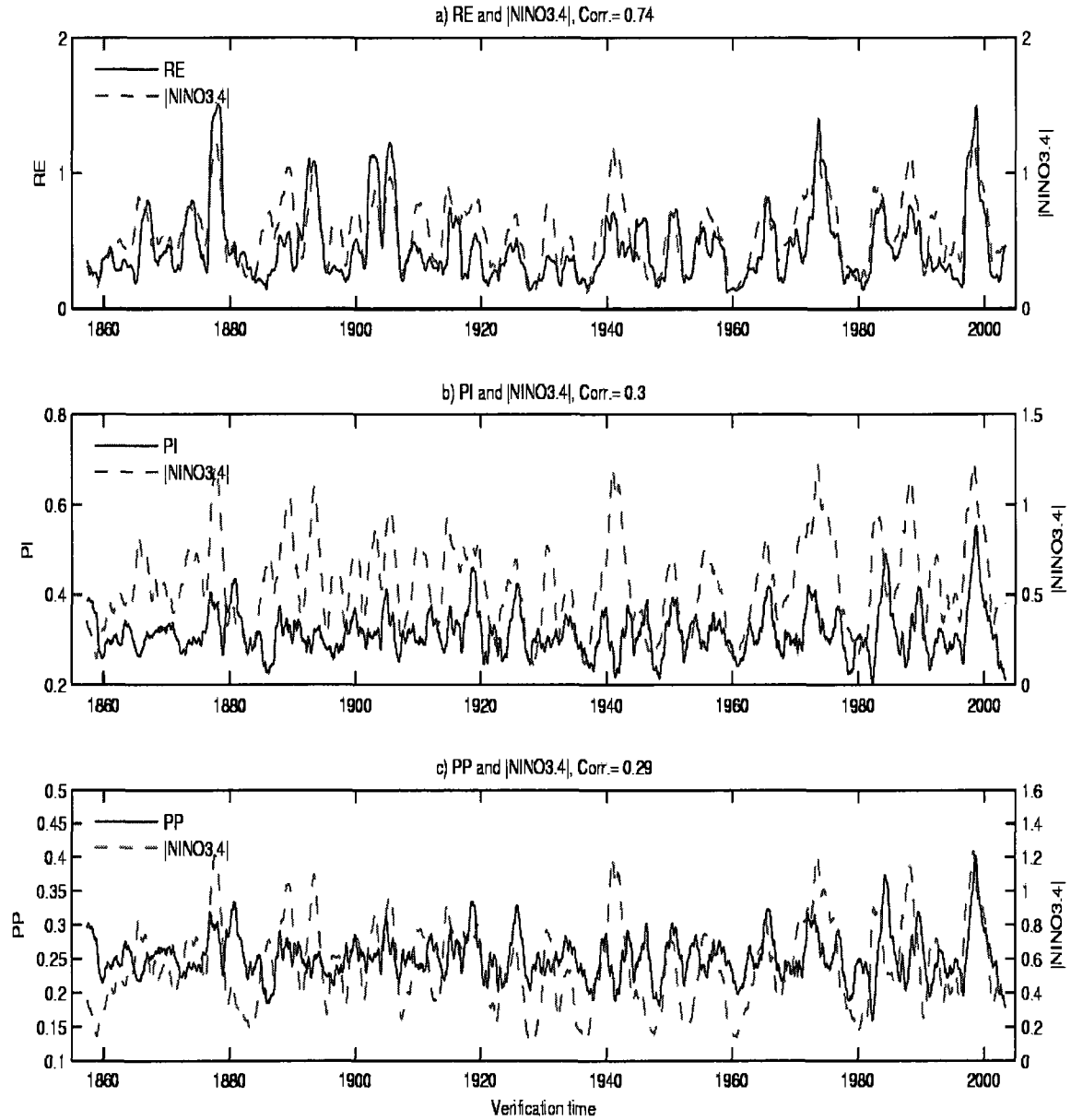


Fig. 5.14. Variations of a) RE, b) PI, and c) PP at the lead time of 6-month along with the time series of absolute value of the NINO3.4 index of SSTA. (A 2-yr running mean method was applied).

Chapter 6: Conclusion and Discussion

This thesis focused on some key issues related to ENSO predictability for the period of 148 yrs (1856-2003) using the LDEO5 model. A number of new findings and conclusions on ENSO predictability were obtained through SV analysis, ensemble prediction, and the relationship between actual predictability and potential predictability, as summarized below.

6.1. Thesis Summary

To study the first kind of predictability (i.e., the loss of predictability due to uncertainty of the initial conditions in the chaotic system), a fully physically-based tangent linear model was constructed for the ZC model, based on which singular vector (SV) analysis was performed for the last 148 years. The results show that the leading SV (SV1) is dominated by a west-east dipole spanning the equatorial Pacific with one center located in the east and the other at the dateline. The spatial distribution of SV1 is less sensitive to initial conditions while the singular values are very sensitive to initial conditions. This property of the SV offers a useful way in ensemble construction and ensemble predictions for the ZC model. On the other hand, the singular value and final pattern represent the amplitude and spatial pattern of error/perturbation growth; they are very sensitive to initial conditions. With this property of final pattern and singular value, further SV analyses were performed to identify the perturbation growth of individual thermodynamical heating processes. Through these SV analyses, dynamical control factors of error/perturbation growth in the LDEO5 model were obtained. It was found that: i) the total linear heating causes a warming effect and controls two positive perturbation growth regions: one in the central Pacific and the other in the

eastern Pacific; whereas the total linearized nonlinear advection brings a dumping effect controlling the negative perturbation growth in the central Pacific; ii) the dynamical diagnosis shows that the total linear and nonlinear heating terms play opposite roles in controlling the optimal perturbation growth, and that the linear optimal perturbation is more than twice as large as the nonlinear one for the ZC model.

In chapter 3, relationships of total perturbation growth rate $S1$, linear perturbation growth rate $L1$ and nonlinear perturbation growth rate $N1$ and actual prediction skills were investigated on multiple time scales from seasonal to decades over the 148 yrs. The $S1$ shows a strong scale-dependent feature with the actual prediction skill which prevents it from being a good indicator of ENSO predictability. On the interdecadal time scale, there is larger model error growth associated with lower model skill and weaker ENSO signals. However, on the shorter time scales, the $S1$ does not have close relations with the actual prediction skills. The fundamental reason for this was revealed by cross-wavelet analysis and mathematical methods as well. There exists a strong offsetting effect in the underlying linear and nonlinear error growth rates. The offsetting effect is represented by $S1 \approx L1 - N1$. The “negative” perturbation growth rate $-N1$ (denoted as $NN1$) probably is the consequence of the unrealistic nonlinear dumping in the LDEO5 model. Although the correlations of the actual prediction skill to both the $L1$ and the $NN1$ are good, their opposite signs lead to a weak relationship between $S1$ and actual prediction skill. Therefore, either $L1$ or $N1/NN1$ is better than $S1$ in measuring actual prediction skill for the LDEO5 model.

A good ensemble construction strategy is required to generate a reliable and high resolution ENSO ensemble-based probabilistic prediction, which is a key issue in studying

ENSO predictability. Furthermore, a reliable ensemble prediction system is an indispensable basis for studying potential predictability. Without a good ensemble construction method, ensemble-based and information-based potential predictability measures would be questionable. Therefore, in chapter 4, ensemble-based probabilistic ENSO predictions were performed with four typical ensemble construction strategies. Several probabilistic prediction verification measures/scores were used to examine the two key properties of our probabilistic predictions, i.e. reliability and resolution. It was found that the stochastic optimal winds contribute to a more reliable and higher resolution probabilistic ENSO prediction, leading to significantly improved RMSE skill and ranked probability score (RPS) at longer lead times. It indicates the important role of atmospheric stochastic forcing in improving ENSO predictability in the LDEO5 model at longer lead times. More models experiments and other construction strategies are needed to further verify this point in future studies.

In Chapter 5, relationships of these potential predictability measures and actual predictability measures were investigated on multiple time scales from interannual to decadal using information theory and ensemble predictions. The object of this chapter is to find reliable and robust potential measures of ENSO predictability from ensemble predictions, which do not make use of observations; instead, these measures rely on the quality of ensemble predictions (i.e. ensemble mean/ensemble mean square, ensemble spread). As aforementioned, these potential predictability measures were obtained based on a reliable and high resolution ensemble construction method (i.e. a joint perturbation method SV1_sst+SO1_winds from chapter 4). It was found that relative entropy (RE) is better than predictive information (PI) and predictive power (PP) in quantifying the correlation-based

prediction skill; whereas PI/PP is a better indicator in estimating mean square error (MSE)-based prediction skill. RE is dominated by the signal component but the dispersion component has comparable contribution during weak ENSO periods. The best potential predictability measures are validated by long-term statistical analyses (i.e. cross-wavelet analyses, FFT filter, running methods); time scale-dependent features were revealed for the relationships of potential and actual predictability measures. We found that the primary reason of the weak relationship between RE and MSE skill is due to the time-scale-dependent property of the RE–MSE relationship, i.e., they have an inverse relationship at interannual time scale but an in-phase relationship at decadal/interdecadal time scale. Similarly, the ensemble ratio (ER) offers the best ensemble-based potential measure to indicate correlation predictability than the EM2 and ES.

In addition, target-time-dependent and spring barrier features were identified using correlation skill, RMSE, ES and PP/PI. Similar features are consistent with those shown in the NCEP CFS model or the LDEO4 model. On the other hand, the RE/ER does not have as strong a target-dependent feature as the other potential predictability measures because it includes both signal and noise components, representing the ratio of signal to noise.

6.2. Discussion

The SV analyses and retrospective hindcasts are often model-dependent, suggesting that the results and conclusion drawn from this work might not be generalized. More models are required to fully generalize these conclusions. It is also necessary to further examine these results and relationships in operational ENSO forecasts.

Recently, another ensemble construction method was shown its ability to generate reliable and high resolution ensemble ENSO predictions based on an intermediate coupled model (ICM) in Zheng et al. (2009). A data assimilation-based ensemble Kalman filter (EnKF) method was applied as a good ensemble construction method for an intermediate coupled model. And a linear, first-order Markov stochastic model is embedded within the SST anomaly model of the ICM to represent the model uncertainties of forecasted SSTA. The improvement of ensemble prediction skills is very significant. Therefore, it may be necessary to test other ensemble construction methods in the ZC model in future studies.

As shown in Chapter 5, the “spring barrier” effect has a large impact on ENSO predictability, but the mechanism of this phenomenon has not been addressed in this study. Considering that the spring barrier is a common issue in many models, it needs to be explored in future work. In our study, the SV is a linear optimal perturbation pattern because it is obtained by SV analysis and with the tangent linear model. On the other hand, there could exist a nonlinear optimal perturbation in ENSO models as proposed by Mu et al. (2003), namely, the conditional optimal perturbation (CNOP). Mu et al. (2007) found that the CNOP-type error has a significant season-dependent evolution, and suggested that the CNOP-type errors can be considered as one of candidate errors that cause the spring

predictability barrier. Xu and Duan (2008) investigated CNOP in the ZC model and suggested that the CNOP yields a larger error growth in strong ENSOs than does a linear SV.

Some physical processes are either simplified or missing in the ZC model. For example, stochastic atmospheric noise is not considered in this model. Stochastic forcing has been thought to be a main source to limit ENSO predictability. Thus, the predictive skill shown in the ZC model might be a lower bound of ENSO actual prediction skill (Chen and Cane 2008). We identified the potential model error from this study: i.e., the total nonlinear heating always contributes to a dumping effect in the ZC model, which is opposite to the observation as discussed in An and Jin (2004), the vertical nonlinear warming should dominate the total nonlinear heating term. This is mainly due to unrealistic model simulation of the zonal current anomaly. Therefore, more accurate parameterizations of these thermodynamical processes in the ZC model will be helpful in improving the predictability.

Nevertheless, this work is the first long-term (>100 years) SV analysis study that explored ENSO predictability, and especially offered a practical means (linear and nonlinear singular values) to estimate the confidence level for individual forecasts. The controlling dynamical factors for individual error growth rates were investigated. Several ensemble construction methods were verified by probabilistic verification methods, from which a reliable and high resolution ensemble construction method was proposed for the ZC model. Finally, through ensemble predictions and with the newly developed information theory, we verified some potential predictability measures i.e. relative entropy and ensemble ratio have close relations with correlation skill; meanwhile, PI/PP has close relation with MSE-based prediction skill measures.

References:

- An S and Jin FF (2004) Nonlinearity and asymmetry of ENSO. *J Clim* 17:2399–2412.
- An S, Ham YG, Kug JS, Jin FF, and Kang IS (2005) El Niño–La Niña Asymmetry in the Coupled Model Intercomparison Project Simulations. *J Clim* 18:2617–2627.
- Anderson JL (1997) Impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low order perfect model results. *Mon Wea Rev* 125:2969–2983.
- Battisti DS (1988) The dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model. *J Atmos Sci* 45(20):2889–2919
- Bishop C, Etherton BJ, and Majumdar S (2001) Adaptive sampling with the Ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon Wea Rev* 129:420–436.
- Bishop CH, and Toth Z (1999) Ensemble transformation and adaptive observations. *J Atmos Sci* 56:1748–1765.
- Blanke B, Neelin JD, and Gutzler D (1997) Estimating the effect of stochastic wind stress forcing on ENSO irregularity. *J Clim* 10:1473–1486.
- Buizza R and Palmer TN (1998) Impact of ensemble size on ensemble prediction. *Mon Wea Rev* 126:2503–2518.
- Buizza R (1997) Potential forecast skill of ensemble prediction and spread and skill distribution of the ECMWF ensemble prediction system. *Mon Wea Rev* 125:99–119.

- Cai M, Kalnay E, Toth Z (2003) Bred vectors of the Zebiak–Cane Model and their potential application to ENSO predictions. *J Clim* 16:40-56
- Chen D, and Cane MA (2008) El Niño prediction and predictability. *J Computational Phys* 227:3625-3640
- Chen D, Cane MA, Kaplan A, Zebiak SE, and Huang D (2004) Predictability of El Niño over the past 148 years. *Nature* 428: 733-736.
- Chen YQ, Battisti DS, Palmer RN, Barsugli J, Sarachik E (1997) A study of the predictability of tropical Pacific SST in a coupled atmosphere/ocean model using singular vector analysis. *Mon Wea Rev* 125:831-845
- DelSole T (2004) Predictability and information theory. Part I: Measures of predictability. *J Atmos Sci* 61:2425–2440.
- DelSole T (2005) Predictability and information theory. Part II: Imperfect forecasts. *J Atmos Sci* 62:3368–3381.
- DelSole T and Tippett MK (2007) Predictability: recent insights from information theory. *Rev Geophys* 45 doi: 10.1029/2006RG000202.
- DelSole T and Tippett MK (2008) Predictable components and singular vectors. *J Atmos Sci* 65:1666-1678.
- Deng Z and Tang Y (2008) The retrospective prediction of ENSO from 1881-2000 by a hybrid coupled model - (II) Interdecadal and decadal variations in predictability. *Clim*

Dyn doi:10.1007/s00382-008-0398-2.

Descamps L and Talagrand O (2007) On some aspects of the definition of initial conditions for ensemble prediction. *Mon Wea Rev* 135: 3260-3272.

Eckert C and Latif M (1997) Predictability of a stochastically forced hybrid coupled model of the tropical Pacific ocean-atmosphere system. *J Clim* 10:1488-1504.

Eisenman I, Yu LS, and Tziperman E (2005) Westerly wind bursts: ENSO's tail rather than the dog? *J Clim* 18: 5224-5238.

Epstein ES (1969) A scoring system for probability forecasts of ranked categories. *J Appl Meteor* 8:985-987.

Evensen G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 43-62.

Evensen G. (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn* 53:343-367

Fan Y, Allen MR, Anderson DLT, Balmaseda MA (2000) How predictability depends on the nature of uncertainty in initial conditions in a coupled model of ENSO. *J Clim* 13:3298-3313.

Farrell BF and Ioannou PJ (1993) Stochastic dynamics of baroclinic waves, *J Atmos Sci* 50:4044-4057.

Fluegel M, Chang P, and Penland C (2004) The role of stochastic forcing in modulating

ENSO predictability. *J Clim* 17:3125–3140.

Gebbie G, Eisenman I, Wittenberg A, and Tziperman E (2007) Modulation of westerly wind bursts by sea surface temperature: A semistochastic feedback for ENSO. *J Atmos Sci* 64: 3281–3295.

Gill AE (1980) Some simple solutions for heat-induced tropical circulation. *Q J R Meteor Soc* 106:447–462

Goddard L, Mason SJ, Zebiak SE, Ropelewski CF, Basher R, and Cane MA (2001) Current approaches to climate prediction. *Int. J. Climatology* 21:1111-1152.

Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlinear Processes in Geophys* 11:561-566.

Ham YG., Kug JS, and Kang IS (2009) Optimal initial perturbations for El Niño ensemble prediction with ensemble Kalman filter. *Clim Dyn*, DOI 10.1007/s00382-009-0582-z.

Hamill TM (1997) Reliability diagrams for multicategory probabilistic forecasts. *Wea Forecasting* 12:736–741.

Hamill TM, Snyder C, and Morss RE (2000) A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon Wea Rev* 128: 1835–1851.

Houtekamer PL and Derome J (1995) Methods for ensemble prediction. *Mon Wea Rev* 123: 2181–2196.

- Jin EK, Kinter JL III, Wang B, Park CK, Kang IS, Kirtman BP, Kug JS, Kumar A, Luo JJ, Schemm J, Shukla J, Yamagata T (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere model. *Clim Dyn* 31:647–664.
- Jin F-F, Lin L, Timmermann A, and Zhao J (2007) Ensemble-mean dynamics of the ENSO recharge oscillator under state-dependent stochastic forcing. *Geophys Res Lett* 34 L03807, doi: 10.1029/2006GL027372.
- Jin F-F, Neelin JD, and Ghil M (1994) El Niño on the devil's staircase: Annual subharmonic steps to chaos. *Science* 264:70–72.
- Karspeck AR, Kaplan A, and Cane MA (2006) Predictability loss in an intermediate ENSO model due to initial error and atmospheric noise, *J Clim* 19(15): 3572-3588.
- Kirtman BP and Schopf PS (1998) Decadal variability in ENSO predictability and prediction. *J Clim* 11: 2804–2822.
- Kirtman BP, Shukla J, Balmaseda M, Graham N, Penland C, Xue Y, Zebiak S (2002) Current status of ENSO forecast skill. A report to the Climate Variability and Predictability. (CLIVAR) Numerical Experimentation Group (NEG), CLIVAR Working Group on Seasonal to Interannual Prediction (Available at http://www.clivar.org/publications/wg_reports/wgsip/nino3/report.htm)
- Kirtman BP (2003) The COLA anomaly coupled model: Ensemble ENSO prediction. *Mon Wea Rev* 131: 2324-2341.
- Kirtman BP and Min D (2009) Multimodel Ensemble ENSO Prediction with CCSM and

- CFS. *Mon. Wea. Rev.*, 137, 2908-2930.
- Kleeman R (2002) Measuring dynamical prediction utility using relative entropy. *J Atmos Sci* 59: 2057–2072.
- Kleeman R (2008) Limits, variability, and general behavior of statistical predictability of the midlatitude atmosphere. *J Atmos Sci* 65:263-275.
- Kleeman R and Moore AM (1997) A theory for the limitation of ENSO predictability due to stochastic atmospheric transients, *J Atmos Sci*, 54: 753–767
- Kumar A, Barnston AB, Peng P, Hoerling MP, and Goddard L (2000) Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. *J Clim* 13: 3139–3151.
- Latif M, Anderson D, Barnett T, Cane M, Kleeman R, Leetma A, O’Brien J, Rosati A, Schneither E (1998) A review of the predictability and prediction of ENSO. *J Geophys Res* 103(C7):14375–14393.
- Leith CE (1974) Theoretical skill of Monte Carlo forecasts. *Mon Wea Rev* 102:409–418.
- Li Z, Navon IM, Hussaini YM (2005) Analysis of the singular vectors of the full-physics FSU global spectral model. *Tellus A* 57:560–574.
- Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus* 17:321-333.
- Lorenz EN (1963) Deterministic nonperiodic flow. *J Atmos Sci* 20:130–141.

- Mason IB (2003) Section 3: Binary Events In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (eds I.T. Jolliffe and D.B. Stephenson). John Wiley & Sons Ltd., England, 137–163 pp.
- Mason SJ, Graham NE (1999) Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea Forecast* 14:713-725.
- Mason S (2004) On using “climatology” as a reference strategy in the Brier and ranked probability skill scores. *Mon Wea Rev* 132:1891-1895.
- Moore AM and Kleeman R (1996) The dynamics of error growth and predictability in a coupled model of ENSO. *Q J R Meteor Soc* 122:1405–1446.
- Moore AM and Kleeman R (1998) Skill assessment for ENSO using ensemble prediction, *Q J R Meteor Soc* 124:557–584.
- Moore AM, Kleeman R (1999) Stochastic forcing of ENSO by the intraseasonal oscillation. *J Clim* 12: 1199-1220.
- Moore AM, Zavala-Garay J, Tang Y, Kleeman R, Weaver AT, Vialard J, Sahami K, Anderson DLT, and Fisher M (2006) Optimal forcing patterns of coupled models. *J Clim* 19: 4683-4699.
- Mu M, Xu H, and Duan W (2007) A kind of initial errors related to “spring predictability barrier” for El Niño events in Zebiak-Cane model. *Geophys Res Lett* 34:L03709, doi:10.1029/2006GL027412.

- Mu M, Duan W, and Wang B (2003) Conditional nonlinear optimal perturbation and its applications. *Nonlinear Proc Geophys* 10:493–501.
- Murphy AH (1969) On the ranked probability skill score. *J Appl Meteor* 8:988-989.
- Murphy AH(1971) A note on the ranked probability skill score. *J Appl Meteor* 10:155-156.
- Murphy AH (1973) A new vector partition of the probability score. *J Appl Meteor* 12: 595-600.
- Palmer TN (2000) Predicting uncertainty in forecasts of weather and climate. *Rep Prog Phys*, 63: 71-116, doi:10.1088/0034-4885/63/2/201
- Palmer TN (1993) Extended-range atmospheric prediction and the Lorenz Model. *Bull Am Meteorol Soc* 74:49–66.
- Peng P and Kumar A (2005) A large ensemble analysis of the influence of tropical SSTs on seasonal atmospheric variability. *J Clim* 15:1068–1085.
- Penland C and Sardeshmukh PD (1995) The optimal growth of tropical sea surface temperature anomalies. *J Clim* 8: 1999–2024.
- Perez CL, Moore AM, Zavaly-Garay J, and Kleeman R (2005) A comparison of the influence of additive and multiplicative stochastic forcing on a coupled model of ENSO. *J Clim* 18:5066-5985.
- Philip SY, and Oldenborgh G.J V (2009) Atmospheric properties of ENSO: models versus observations. *Clim Dyn* doi:10.1007/s00382-009-0579-7.

- Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool H M, Pan HL, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, and Xie P (2006) The NCEP climate forecast system. *J Clim* 19: 3483-3517.
- Scherrer, SC, Appenzeller C, Eckert P, and Cattani D (2004) Analysis of the spread-skill relations using the ECMWF ensemble prediction system over Europe. *Wea Forecast* 19: 552-565.
- Schneider T and Griffies SM (1999) A conceptual framework for predictability studies. *J Clim* 12:3133-3155.
- Shukla J (1998) Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science* 282:728-731
- Stephenson DB and Doblas-Reyes FJ (2000) Statistical methods for interpreting Monte Carlo forecasts. *Tellus* 52A:300-322.
- Stern W and Miyakoda K (1995) Feasibility of seasonal forecasts inferred from multiple GCM simulations. *J Clim* 8:1071-1085
- Suarez MJ and Schopf PS (1988) A delayed action oscillator for ENSO. *J Atmos Sci* 45: 3283-3287
- Tang Y and Deng Z (2009) Low-dimensional nonlinearity of ENSO and its impact on predictability. *Physical D*. doi:10.1016/j.physd.2009.11.006.
- Tang Y, Deng Z, Zhou X, Cheng Y, and Chen D (2008c) Interdecadal Variation of ENSO

Predictability in Multiple Models. *J. Clim* 21: 4811-4833.

Tang Y, Hsieh WW (2003) ENSO simulation and predictions using a hybrid coupled model with data assimilation. *J Japan Meteor Soc* 81(1):1-19

Tang Y, Kleeman R, and Miller S (2006) ENSO predictability of a fully coupled GCM model using singular vector analysis. *J Clim* 19:3361-3377.

Tang Y, Kleeman R, and Moore A (2008a) Comparison of Information-based Measures of Forecast Uncertainty in Ensemble ENSO Prediction. *J Clim* 21 (2): 230-247.

Tang Y, Lin H, and Moore A (2008b) Measuring the potential predictability of ensemble climate predictions. *J Geophys Res* 113. D04108, doi: 10.1029/2007JD008804

Tang Y, Kleeman R, and Moore A (2005) On the reliability of ENSO dynamical predictions. *J Atmos Sci* 62(6): 1770–1791.

Tang Y, Lin H, Derome J, and Tippett MK (2007) A predictability measure applied to seasonal predictions of the Arctic Oscillation. *J. Clim* 20: 4733-4750.

Thompson CJ and Battisti DS (2000) A linear stochastic dynamical model of ENSO. Part I: Model development. *J Clim* 13:2818–2883.

Tippett MK, Kleeman R, and Tang Y (2004) Measuring the potential utility of seasonal climate predictions. *Geophys. Res.Lett.* 31 L22201,doi:10.1029/2004GL021575.

Torrence C and Compo G.P (1998) A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79 (1): 61–78.

- Toth Z and Kalnay E (1993) Ensemble forecasting at NMC: The generation of perturbations. *Bull Am Meteor Soc* **74**:2317–2330.
- Toth Z and Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon Wea Rev* **125**: 3297–3319.
- Toth Z, Talagrand O, Candille G, and Zhu Y (2003) Probability and ensemble forecasts. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (eds I.T. Jolliffe and D.B. Stephenson). John Wiley & Sons Ltd., England, 137–163.
- Tziperman E, and Yu L (2007) Quantifying the dependence of westerly wind bursts on the large scale tropical Pacific SST. *J Clim* **20**:2760-2768.
- Vecchi G.A and Harrison D.E (2003) On the termination of the 2002–03 El Niño event. *Geophys Res Lett* **30**:1946, doi:10.1029/2003GL017564
- Vialard J, Vitart F, Balmaseda M.A, Stockdale T.N, and Anderson D.L (2005) An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. *Mon Wea Rev* **133**: 441-453
- Wang B (1995) Interdecadal changes in El Niño onset in the last four decades. *J Clim* **8**: 267–285.
- Wang X and Bishop C (2003) A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J Atmos Sci* **60**:1140–1158.
- Wei M, Toth Z, Wobus R, and Zhu Y (2008) Initial perturbations based on the ensemble

transform (ET) technique in the NCEP global ensemble forecast systems. *Tellus* 60A: 62–79.

Wilks DS (1995) *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series. Vol. 59, Academic Press, 467 pp.

Xu H and Duan WS (2008) What Kind of Initial Errors Cause the Severest Prediction Uncertainty of El Niño in Zebiak-Cane model. *Adv Atmos Sci* 25: 577–584. DOI:10.1007/s00376-008-0577-4.

Xue Y, Cane MA, Zebiak SE (1997a) Predictability of a coupled model of ENSO using singular vector analysis, Part I: optimal growth in seasonal background and ENSO cycles. *Mon Wea Rev* 125:2043–2056.

Xue Y, Cane MA, Zebiak SE (1997b) Predictability of a coupled model of ENSO using singular vector analysis, Part II: Optimal growth and forecast skill. *Mon Wea Rev* 125:2057-2073

Zavala-Garay, Zhang JC, Moore A, and Kleeman R (2005) The linear response of ENSO to the Madden-Julian Oscillation. *J Clim* 18: 2441-2459.

Zebiak SE and Cane MA (1987) A model El Niño-Southern Oscillation. *Mon Wea Rev* 115: 2262–2278.

Zhang RH and Busalacchi AJ (2008) Rectified effects of tropical instability wave (TIW)-induced atmospheric wind feedback in the tropical Pacific. *Geophys Res Lett* 35: L05608, doi:10.1029/2007GL033028.

Zheng F, Zhu J, Wang H, and Zhang RH (2009) Ensemble hindcasts of ENSO events over the past 120 years using a large number of ensembles. *Adv Atmos Sci* 26(2):359-372. doi:10.1007/s00376-009-0359-7.

Zhou X, Tang Y, and Deng Z (2008) The impact of nonlinear atmosphere on the fastest error growth of ENSO prediction. *Clim Dyn* 30 (5):519-531. DOI: 10.1007/s00382-007-0302-5.